

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200580031943. X

[51] Int. Cl.

C12N 15/861 (2006.01)

C12N 15/63 (2006.01)

C12N 15/64 (2006.01)

C12N 15/10 (2006.01)

A61K 48/00 (2006.01)

C12N 1/10 (2006.01)

[43] 公开日 2007年9月12日

[11] 公开号 CN 101035901A

[51] Int. Cl. (续)

C07H 21/04 (2006.01)

C12Q 1/68 (2006.01)

C12Q 1/70 (2006.01)

[22] 申请日 2005.7.21

[21] 申请号 200580031943. X

[30] 优先权

[32] 2004. 7. 22 [33] US [31] 60/589,804

[32] 2005. 3. 11 [33] US [31] 60/660,542

[86] 国际申请 PCT/US2005/025833 2005.7.21

[87] 国际公布 WO2006/012393 英 2006.2.2

[85] 进入国家阶段日期 2007.3.22

[71] 申请人 细胞基因系统有限公司

地址 美国加利福尼亚州

[72] 发明人 琳达·霍金 雷迪·鲍里斯·塞西达

[74] 专利代理机构 北京金之桥知识产权代理有限公司

代理人 梁朝玉

权利要求书4页 说明书220页 附图12页

[54] 发明名称

腺病毒载体中转基因的插入

[57] 摘要

提供一种腺病毒载体，其含有插入在病毒基因组新位点的转基因，其使得一个或多个天然加工、调节信号或诸如分支点位点、剪接受体位点、内部核糖体进入位点、自加工切割位点和/或 poly A 信号的异源转录信号有助于转基因表达。

1. 一种包含转基因编码序列（CDS）的腺病毒载体，其中由嵌合腺病毒前导区表达所述 CDS，其中嵌合腺病毒前导区的天然部分选自 L1、L2、L3 和 L4 腺病毒前导区。
2. 如权利要求 1 所述的腺病毒载体，其中所述嵌合腺病毒前导区包含分支点和剪接受体位点、转基因 CDS 和腺病毒 CDS。
3. 如权利要求 1 所述的腺病毒载体，其中嵌合腺病毒前导区包含第一腺病毒 CDS、分支点、第一剪接受体位点、转基因 CDS、第二分支点、第二剪接受体位点和第二腺病毒 CDS。
4. 如权利要求 1 所述的腺病毒载体，其中嵌合腺病毒前导区依次包含腺病毒 CDS、分支点、剪接受体位点、转基因 CDS 和多腺苷酸化信号。
5. 如权利要求 1 所述的腺病毒载体，其中嵌合腺病毒前导区依次包含腺病毒 CDS、可操作地连接转基因 CDS 的内部核糖体进入位点和多腺苷酸化信号。
6. 如权利要求 2 所述的腺病毒载体，其中第一剪接受体位点是内源腺病毒剪接受体位点或异源剪接受体位点。
7. 如权利要求 2 所述的腺病毒载体，其中第二剪接受体位点是内源腺病毒剪接受体位点或异源剪接受体位点。
8. 如权利要求 2 所述的腺病毒载体，其中第一分支点是内源腺病毒分支点或异源分支点。
9. 如权利要求 2 所述的腺病毒载体，其中第二分支点是内源腺病毒分支点或异源分支点。
10. 如权利要求 4 所述的腺病毒载体，其中剪接受体位点是异源剪接受体位点。
11. 如权利要求 4 所述的腺病毒载体，其中分支点是异源分支点。
12. 如权利要求 1 所述的腺病毒载体，其中嵌合腺病毒前导区依次包含可操作地连接自加工切割位点的腺病毒 CDS 和可操作地连接所述自加工切割位点的转基因 CDS。

13. 如权利要求 1 所述的腺病毒载体, 其中嵌合腺病毒前导区依次包含可操作地连接自加工切割位点 CDS 的转基因 CDS 和可操作地连接所述自加工切割位点的腺病毒 CDS。

14. 如权利要求 12 所述的腺病毒载体, 其中所述自加工切割位点 CDS 是 2A 或 2A-样序列。

15. 如权利要求 12 所述的腺病毒载体, 其中所述自加工切割位点 CDS 选自 SEQ ID NO: 1-32。

16. 如权利要求 1 所述的腺病毒载体, 其中嵌合腺病毒前导区的天然部分来自 L2 腺病毒前导区。

17. 如权利要求 16 所述的腺病毒载体, 其中嵌合腺病毒前导区依次包含 pVII CDS、第一剪接受体位点、转基因 CDS、第二剪接受体位点和 pV CDS。

18. 如权利要求 17 所述的腺病毒载体, 其中所述第一剪接受体位点是内源腺病毒剪接位点, 所述第二剪接受体位点是异源剪接位点。

19. 如权利要求 17 所述的腺病毒载体, 其中所述第一剪接受体位点是异源剪接位点, 所述第二剪接受体位点是内源腺病毒剪接位点。

20. 如权利要求 16 所述的腺病毒载体, 其中嵌合腺病毒前导区依次包含五邻体 CDS、第一剪接受体位点、转基因 CDS、第二剪接受体位点和 pVII CDS。

21. 如权利要求 20 所述的腺病毒载体, 其中所述第一剪接受体位点是内源腺病毒剪接位点, 所述第二剪接受体位点异源剪接位点。

22. 如权利要求 20 所述的腺病毒载体, 其中所述第一剪接受体位点是异源剪接位点, 所述第二剪接受体位点是内源腺病毒剪接位点。

23. 如权利要求 16 所述的腺病毒载体, 其中嵌合腺病毒前导区依次包含 Mu CDS、剪接受体位点、转基因 CDS 和多腺苷酸化信号。

24. 如权利要求 23 所述的腺病毒载体, 其中所述多腺苷酸化信号是腺病毒 L2 多腺苷酸化信号或异源多腺苷酸化信号。

25. 如权利要求 16 所述的腺病毒载体, 其中嵌合腺病毒前导区依次包含 Mu CDS、内部核糖体进入位点、转基因 CDS 和多腺苷酸化信号。

26. 如权利要求 25 所述的腺病毒载体，其中所述多腺苷酸化信号是腺病毒 L2 多腺苷酸化信号或异源多腺苷酸化信号。

27. 如权利要求 1 所述的腺病毒载体，其中嵌合腺病毒前导区的天然部分来自 L3 腺病毒前导区。

28. 如权利要求 27 所述的腺病毒载体，其中嵌合腺病毒前导区依次包含 23K CDS、剪接受体位点、转基因 CDS 和多腺苷酸化信号。

29. 如权利要求 28 所述的腺病毒载体，其中所述多腺苷酸化信号是腺病毒 L3 多腺苷酸化信号或异源多腺苷酸化信号。

30. 如权利要求 29 所述的腺病毒载体，其中所述转基因 CDS 编码人肿瘤坏死因子相关凋亡诱导配体蛋白。

31. 如权利要求 27 所述的腺病毒载体，其中嵌合腺病毒前导区依次包含六邻体 CDS、第一剪接受体位点、转基因 CDS、第二剪接受体位点和 23K CDS。

32. 如权利要求 31 所述的腺病毒载体，其中所述第一剪接受体位点是内源腺病毒剪接位点，所述第二剪接受体位点是异源剪接位点。

33. 如权利要求 31 所述的腺病毒载体，其中所述第一剪接受体位点是异源剪接位点，所述第二剪接受体位点是内源腺病毒剪接位点。

34. 如权利要求 27 所述的腺病毒载体，其中嵌合腺病毒前导区依次包含 23K CDS、内部核糖体进入位点、转基因 CDS 和多腺苷酸化信号。

35. 如权利要求 35 所述的腺病毒载体，其中所述转基因编码人肿瘤坏死因子相关凋亡诱导配体蛋白。

36. 一种包含转基因编码序列（CDS）的腺病毒载体，其中由嵌合腺病毒前导区表达所述转基因 CDS，其中嵌合腺病毒前导区的天然部分来自 L5 区，嵌合腺病毒区依次包含第一剪接受体位点、转基因 CDS、第二剪接受体位点、尾丝 CDS 和多腺苷酸化信号。

37. 如权利要求 36 所述的腺病毒载体，其中所述第一剪接受体位点是内源腺病毒剪接位点，所述第二剪接受体位点是异源剪接位点。

38. 如权利要求 36 所述的腺病毒载体，其中所述第一剪接受体位点是异源剪接位点，所述第二剪接受体位点是内源腺病毒剪接位点。

39. 一种包含转基因编码序列（CDS）的腺病毒载体，其中由嵌合腺病毒前导区表达所述转基因 CDS，其中嵌合腺病毒前导区的天然部分来自 L5 区，嵌合腺病毒区依次包含第一剪接受体位点、尾丝 CDS、第二剪接受体位点、转基因 CDS 和多腺苷酸化信号。

40. 如权利要求 39 所述的腺病毒载体，其中所述第一剪接受体位点是内源腺病毒剪接位点，所述第二剪接受体位点是异源剪接位点。

41. 如权利要求 39 所述的腺病毒载体，其中所述第一剪接受体位点是异源剪接位点，以及所述第二剪接受体位点是内源腺病毒剪接位点。

42. 如 权利要求 39 所述的腺病毒载体，其中所述多腺苷酸化信号是 L5 多腺苷酸化信号。

腺病毒载体中转基因的插入

本申请要求 2004 年 7 月 22 日提交的序列号为 60/589,804 的美国临时申请和 2005 年 3 月 11 日提交的序列号为 60/660,542 的美国临时申请的优先权。

技术领域：

本发明涉及包含在基因组不同位置插入转基因的，使得转基因表达依赖于天然腺病毒加工和/或调节信号或异源转录信号的腺病毒载体及其制备方法。

背景技术：

在哺乳动物细胞中，腺病毒载体已广泛地用于体外、离体和体内转移并表达基因。腺病毒载体为基因转移提供了许多有利条件：在健康个体中野生型腺病毒仅引起轻微的疾病，其易于增殖到高滴度，其能感染大多数的细胞类型且与它们的生长状态无关，其能将感兴趣的编码蛋白的核酸转移至细胞，且在其最近的具体实践中，其能被改造以在特定细胞中选择性地复制，并能通过结合或依附衣壳蛋白配位体而被靶向。

存在着两种主要类型的腺病毒载体，非复制型（复制缺陷型）和复制型。传统上，复制缺陷型载体缺少复制所必需的一种或多种基因。这些非复制型病毒在具有其所缺少的必需基因的细胞中增殖。传统上，复制型腺病毒载体并不缺少复制所必需的任何腺病毒基因。一种类型的复制型腺病毒在特定的细胞中能选择性地复制。例如，已描述了在特定细胞类型中优先复制（并因此破坏这些特定细胞）的多种不同的细胞和组织特异性复制型腺病毒载体。一种构建细胞特异性复制型病毒的方法是与必需的腺病毒编码序列可操作地连接的方式来放置异源转录调节元件（TRE）。另一种方法是当所要被缺失或失活的腺病毒编码序列在某种细胞类型如肿瘤细胞中能引起病毒优先复制时，缺失或失活该序列。

腺病毒载体受限于其基因组尺寸（Bett 等, *J Virol* 67:5911 - 5921, 1993）。这又限制了能插入载体的异源 DNA 数量, 并因此限制了能插入在腺病毒基因组的异源编码序列数量和/或长度。因此, 复制缺陷型病毒能插入最大的异源 DNA, 但仍受限于腺病毒基因组 DNA 尺寸以及腺病毒 DNA 缺失的数量。

至于复制型病毒, 它们受限不仅因为它们包含大部分腺病毒编码序列, 还在于异源 DNA 插入位点的数量。传统上, 已插入 DNA 异源插入物以取代复制型病毒中腺病毒 E3 编码序列（WO 02/067861, WO 01/02540）。这限制了异源 DNA 插入物的尺寸, 仅稍大于 E3 缺失。同样地, 对于特定的应用, 缺失 E3 编码序列或在 E3 区中插入异源 DNA 并非所期望或有利的。

根据病毒 DNA 复制, 腺病毒感染分为早期和晚期。在病毒 DNA 复制前（早期）和/或在病毒 DNA 复制后（晚期）病毒基因被表达。一般而言, 早期基因编码细胞为病毒增殖所准备的蛋白, 而晚期蛋白是衣壳蛋白或与病毒颗粒包装和装配有关。一种已广泛使用的表达转基因的方法是通过病毒自身组成型及非自调节型表达的方式将盒插入在病毒基因组中。在这种情况下, 一般地转基因插入取代了非必需基因或必需基因, 其中必需基因为用于病毒载体增殖的细胞系所补充。

在腺病毒中, DNA 复制后主要由主要晚期启动子（MLP）启动晚期转录。在晚期时间点, MLP 启动的转录占总 RNA 合成的大约 30%。在所有腺病毒中, 初级转录物从 MLP 开始向右延伸约 28 kb 并在接近于病毒基因组 99 图距单位的位置终止。每个区都具有自己的多腺苷酸化信号序列, 通常由几个不同的剪接信使组成。沿着基因组在五个位置中的一个, 初级转录物被切割并多腺苷酸化, 产生五个家族（L1-L5）的 mRNAs, 每个都具有 3'共有末端。在 3'末端选择后, 通过这样一种方式剪接加工初级转录物, 即每个成熟的 RNA 都获得一组共有的源自 16.8、19.8 和 26.9 图距单位的三条短的 5'前导区区段, 称为三联前导序列（TPL）。在某些腺病毒血清型中, 除 L5 之外的五个家族中的每一个都表达超过一种的蛋白。Ad 40 和 41 表达超过一种的来自 L5 区的蛋白。由于在基因组晚期区中, 三联前导序列剪接入开放阅读框（ORF）上游剪接受体位点的位置不

同，因此家族中 mRNAs 之间彼此不同。例如，四种产生自 L2 区的 mRNAs，都具有共有的 3'末端。天然 Ad 转录物中尽管信使是多顺反子的，但只有位于 5'末端的 ORF 被认为将被翻译。Fuerer 等（Gene Therapy 11: 142-151（2004））描述了使用至少两种方法将胞嘧啶脱氨酶（CD）编码序列插入复制型病毒中。一种方法依赖于使用脑炎心肌病毒内部核糖体进入位点（IRES），将 L5 转录物转变为双顺反子 mRNA。Fuerer 等还描述了一种利用源自 Ad41 长尾丝基因的剪接受体位点序列以在主要晚期转录物的三联前导序列外显子上剪接 CD 盒。

发明内容：

本发明涉及一种新的基因转移系统，其使用腺病毒（Ad）基因组作为基因转移和/或表达载体。腺病毒基因组可以是复制型或非复制型（复制缺陷型）。

本发明提供了腺病毒载体，其中转基因以有效地利用和/或使用天然腺病毒加工的位置和方式，插入在腺病毒基因组的不同位置（例如现有基因上游或下游），而无需插入大的调节序列如启动子，以使得异源 DNA（转基因）插入 Ad 基因组。

在本发明某些方面，转录信号（如分支点位点、剪接受体位点、IRES、自加工切割位点和/或 polyA 信号）被添加到本发明的腺病毒载体。具有诸如分支点序列的加工信号的插入基因与剪接受体序列一起利用了天然腺病毒所使用的 mRNA 加工和表达系统。

附图说明：

图 1 是 Ad5 天然基因组结构的示意图。

图 2 是天然 Ad5 转录单位的示意图。

图 3 是 Ad5 L1 区的示意图。

图 4 是 Ad5 L2 区的示意图。

图 5 是 Ad5 L3 区的示意图。

图 6 是 E3 转录单位的示意图。

图 7 是 E3A 区的示意图。星号表示 Ad5 中的 192 碱基对被经常从腺病毒载体中删除。

图 8 是 E3B 区的示意图。

图 9 是本发明一个示范性实施例的示意图，其中 L3 区插入位点显示具有嵌合腺病毒前导区的一个例子，其中转基因插入在六邻体(hexon)和 23K 蛋白酶编码序列之间。转基因可操作地连接（即用于表达）位于 Ad5 中六邻体终止密码子上游约 95 个碱基的用于 23K mRNA 的天然剪接受体位点。异源剪接受体位点是被插入在转基因编码序列和 23K 编码序列之间，并可操作地连接 23K 编码序列。这使得四种不同的 mRNAs 从 L3 区转录。在该例子中，天然 L3 多腺苷酸化序列保持完整，并为嵌合 L3 区的所有四种 mRNAs 所利用。

图 10 是本发明一个示范性实施例的示意图，其中 L3 区插入位点显示具有嵌合腺病毒前导区的一个例子，该情况下，转基因插入在 L3 区中 23K 蛋白酶编码序列和 L3 多腺苷酸化序列之间。转基因可操作地连接（即用于表达）2A-样序列或 IRES。2A-样序列或 IRES 同样可操作地连接 23K 编码序列。类似于天然的 L3 区，这个嵌合腺病毒 L3 区将产生三种 L3 mRNAs。在该情况下，所有三种 L3 mRNAs 都包含 IRES（或 2A-样序列），因此所有三种这些 mRNAs 的翻译将导致转基因的表达。就该例子而言，天然 L3 多腺苷酸化序列保持完整，并为嵌合 L3 区的所有三种 mRNAs 所利用。

图 11 提供了 OV-1160 的示意图，其在 5'-3'方向上包含：可操作地连接 E1A 的 E2F-1 启动子；以及在 L3 区中的：pVI、六邻体、剪接受体和肿瘤坏死因子相关凋亡诱导配体(TRAIL)编码序列。该载体在天然位置携带包装信号，并在 E2F-1 启动子上游携带多腺苷酸化信号以抑制由 LITR 的转录连读。

图 12 提供了 OV-1164 的示意图，在 5'-3'方向上包含：可操作地连接 E1A 的 E2F-1 启动子；以及在 L3 区中的：pVI、六邻体、IRES、TRAIL 编码序列和 23K 基因。该载体在天然位置携带包装信号，并在 E2F-1 启动子上游携带多腺苷酸化信号以抑制由 LITR 的转录连读。

具体实施方式：

本发明提供了用于表达一种或多种转基因的，依赖于使用现有的（天然的）转录和/或翻译调节元件的重组腺病毒载体。在一个例子中，通过插入一种或多种转基因至晚期区，在感染晚期主要晚期启动子用于表达转基因。在另一个例子中，使用该策略将转基因插入在腺病毒早期区中。只要必需腺病毒基因及诸如内源多腺苷酸化和转录信号的天然腺病毒加工信号的开放阅读框（ORF）不被扰乱，并且病毒能有效地复制，转基因就可以可操作地连接任何内源腺病毒启动子的方式插入。

除转基因之外，分支点、剪接受体位点或其它类似序列皆可包含在本发明的腺病毒载体中，以提供对转录物的正确加工。通过改变特定的加工信号序列，可以操纵转基因表达的数量和动力学，例如剪接效率、翻译启动效率或信号效率，以便获得期望的动力学。多个转基因可包含在同一个载体的相同或不同的基因组区中。其它内源腺病毒基因可以被删除，以获得期望的效果或供包含更长的外源 DNA 序列之用。

在一方面，本发明“恢复转录信号”到腺病毒载体。就“恢复转录信号”意味着非天然转录信号被用于执行与天然转录信号类似功能或以类似于天然转录信号的方式影响转录。例如，在 L5 区中尾丝终止密码子包含在 L5 polyA 信号中。在尾丝编码区下游转基因的插入通常包括破坏天然 L5 polyA 位点功能的突变，但仍保留尾丝终止密码子。因此，为了正确加工 L5 信使应恢复 polyA 位点。这需要添加 polyA 信号到 L5 区。该 polyA 信号可以是与被破坏序列相同的序列，或可以是非天然或异源 polyA 信号。在一个转基因被插入尾丝编码区下游的示范性实施例中，polyA 信号被插入转基因编码区下游。在本发明转基因插入在现有腺病毒编码序列之间的另一实施例中，额外的剪接受体位点的恢复/插入通常得到使用。例如，转基因可使用对于转基因编码序列下游邻接的腺病毒编码序列为天然剪接位点的腺病毒天然剪接位点。在该情况下，为了转基因编码序列下游邻接的腺病毒编码序列，该剪接位点必须恢复。该“恢复的”剪接位点可以是与在此用于剪接转基因序列上游的天然序列相同的序列，或可以是异源剪接

位点。在本领域中已知不同的剪接位点具有不同的剪接效率。因此，视具体应用所需，本领域技术人员可选择具有不同剪接效率的剪接位点。

本发明所提供的有利之处在于以下几个方面：a) 通过利用现有表达信号，可在尺寸受限的载体中接入更大尺寸的转基因；b) 已知和可预测的腺病毒基因表达模式使设计具有期望的转基因表达动力学例如表达水平和时限的腺病毒载体成为可能，其中表达是基于转基因的特性而定制；c) 本发明的腺病毒载体利用了已置于腺病毒基因组中的天然调节系统；d) 本发明的腺病毒载体包括导入的调节序列，例如自加工切割位点、分支点序列和剪接受体序列，与其它通常用于转基因表达的元件相比它们是小；和 e) 异源转录因子的选择可用于进一步调节表达。诸多转基因和/或可作为转基因候选物的现有转基因物类，在此得到进一步描述。

除非另有说明，本发明实施应使用处于本领域技术范畴之内的常规分子生物学（包括重组技术）、微生物学、细胞生物学、生物化学和免疫学技术。在文献中详细描述上述技术，例如“Molecular Cloning: A Laboratory Manual”，第2版（Sambrook等，1989）；“Oligonucleotide Synthesis”（MJ. Gait编，1984）；“Animal Cell Culture”（R.I. Freshney编，1987）；“Methods in Enzymology”（Academic Press, Inc.）；“Handbook of Experimental Immunology”（D.M. Weir & CC. Blackwell,编）；“Gene Transfer Vectors for Mammalian Cells”（J.M. Miller & M.P. Calos编，1987）；“Current Protocols in Molecular Biology”（F.M. Ausubel等编，1987）；“PCR: The Polymerase Chain Reaction”，（Mullis等编，1994）；和“Current Protocols in Immunology”（J.E. Coligan等编，1991）。

定义

除非另有说明，所有本文使用的术语具有与本领域技术人员所理解的含义，且本发明实施应使用处于本领域技术人员知识范畴之内的常规微生物学和重组DNA技术。

术语“病毒”、“病毒颗粒”、“载体颗粒”、“病毒载体颗粒”和“病毒粒子”可交换使用，并可宽泛地理解为这样一种含义，即指当例如本发明的病毒载体被导入适当细胞或细胞系中用于产生感染性颗粒时，所形成的感

染性病毒颗粒。为了在体外或体内将 DNA 转移到细胞，可使用根据本发明的病毒颗粒。对本发明来说，这些术语指腺病毒，包括当本发明的腺病毒载体包在腺病毒衣壳中所形成的重组腺病毒。

“腺病毒载体 (adenovirus vector)”或“腺病毒载体 (adenoviral vector)”(可交换使用)在此指一种复制型或非复制型(复制缺陷型)的多核苷酸构建体。

本发明的示范性腺病毒载体包括但不限于，DNA、包在腺病毒包衣中的 DNA、包装在其它病毒或病毒样形式(例如单纯疱疹和 AAV)中的腺病毒 DNA、包在脂质体中的腺病毒 DNA、与多聚赖氨酸复合的腺病毒 DNA、与合成的聚阳离子分子复合的腺病毒 DNA、与铁传递蛋白缀合的腺病毒、或与诸如 PEG 的化合物复合以在免疫学上“掩蔽”抗原性和/或增加半衰期的腺病毒 DNA、或与非病毒蛋白缀合的腺病毒 DNA。因此，本文使用的术语“腺病毒载体”或“腺病毒载体”包括腺病毒或腺病毒颗粒。

术语“多核苷酸”和“核酸”在此可交换使用，指任何长度的核糖核苷酸或脱氧核糖核苷酸的核苷酸多聚形式。这些术语包括单链、双链或三链 DNA，基因组 DNA，cDNA，RNA，DNA-RNA 杂交体，或包含嘌呤和嘧啶碱基或其它天然、化学、生物化学修饰的、非天然或衍生的核苷酸碱基的聚合物。优选地，本发明的载体包含 DNA。本文使用的“DNA”不但包括碱基 A、T、C 和 G，还包括它们的任何类似物或这些碱基的修饰形式，例如甲基化核苷酸、诸如不带电键合和硫代的核苷酸间修饰、糖类似物的使用、以及修饰的和/或可选择的主链结构，例如聚酰胺。

以下是多核苷酸的非限制性例子：基因或基因片段、外显子、内含子、mRNA、tRNA、rRNA、核酶、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、分离的任何序列的 DNA、分离的任何序列的 RNA、核酸探针以及引物。多核苷酸可包含修饰的核苷酸，例如甲基化核苷酸和核苷酸类似物、尿嘧啶、其它糖类以及诸如氟代核糖和硫代的连接基团，以及核苷酸分支。核苷酸序列可为非核苷酸组分所中断。在聚合后多核苷酸可得到进一步修饰，例如通过与标记组分缀合。包括在该定义中的其它类型修饰是加帽、用类似物取代一个或多个天然存在的多核苷酸、以及导入用于连接多核苷酸与蛋白、金属离子、标记组分、其它多核苷酸或固体载

体的试剂。优选地，多核苷酸是 DNA。本文使用的“DNA”不但包括碱基 A、T、C 和 G，而且包括任何一种它们的类似物或这些碱基的修饰形式，例如甲基化核苷酸、诸如不带电键合和硫代的核苷酸间修饰、糖类似物的使用、以及修饰的和/或可选择的骨干结构，例如聚酰胺。当其与其它核酸序列处于一种功能关系时，核酸序列是“可操作性地连接的（operatively linked）”或“可操作地连接的（operably linked）”。例如，如果这两个序列处于使得启动子或调节 DNA 序列影响编码或结构 DNA 序列表达水平时，则启动子或调节 DNA 序列“可操作地连接”或“可操作地连接”到用于编码 RNA 和/或蛋白的 DNA 序列。可操作地连接意味着被连接的 DNA 序列通常是邻接的，在需要连接两个蛋白编码区处，编码区邻接且处于同一个阅读框。然而，因为增强子通常在与启动子相隔几千碱基对时起作用，且内含子序列长度可变，因此某些多核苷酸元件可以是可操作地连接但不邻接。

本文使用的术语“编码区”指含有编码序列的核酸区。编码区可含有其它来自相应基因的区，包括内含子。术语“编码序列”（CDS）指编码蛋白的含有密码子的核酸序列。编码序列通常起始于翻译起始密码子（如 ATG）并终止于翻译终止密码子。在编码序列上游的序列是翻译起始密码子的 5'，在 CDS 下游的序列是翻译终止密码子的 3'。

术语“ORF”指开放阅读框。

术语“基因”指位于基因组内的一段规定的区域，除前述编码序列之外，包含其它负责控制表达即编码部分的转录和翻译的核酸序列，主要是调节序列。基因还可包含其它的 5'和 3'非翻译序列和终止序列。取决于基因源，可存在更多的元件，例如内含子。

本文使用的术语“异源的”和“外源的”是指诸如启动子和基因编码序列的核酸分子，其为对于特定病毒或宿主细胞而言源于外源（非天然）的序列，或如果来自同源，则是修饰自其原型的。因此，病毒或细胞中的异源基因包括对于特定病毒或细胞是内源的，但已通过例如密码子优化而被修饰的基因。该术语还包括非天然存在的天然存在的核酸序列的多拷贝。因此，该术语所指的核酸片段是对于病毒或细胞而言为外源或异源，或对于病毒或细胞是同源的但处于并不常见于宿主病毒或细胞基因组中的位置。

术语“天然的”指存在于野生型病毒或细胞的基因组中的基因。

术语“天然存在的”或“野生型”用于描述可存在于自然界中，不同于人工制造的物体。例如，存在于生物体（包括病毒）中的蛋白或核苷酸序列，其能从天然来源中分离且不被人在实验室中所有意地修饰，是天然存在的。

术语“互补”和“互补的”指两条包含反平行核苷酸序列的核苷酸序列，可以在反平行核苷酸序列互补碱基残基之间形成氢键而彼此配对。至于核酸分子，本文使用的术语“重组的”指使用重组 DNA 技术将核酸分子连接在一起成为后代核酸分子的组合。至于病毒、细胞和生物体，本文使用的术语“重组的”、“转化的”和“转基因的”指其中已导入异源核酸的宿主病毒、细胞或生物体。核酸分子可稳定地整合入宿主基因组中，或者核酸分子还可作为染色体外分子存在。这样的染色体外分子可自发复制。重组病毒、细胞和生物体理解为不仅包含转化加工的终产物，还包含其重组后代。“非转化的”、“非转基因的”或“非重组的”宿主指不含有异源核酸分子的野生型病毒、细胞或生物体。

“调节元件”是与控制核苷酸序列表达有关的序列。调节元件包括启动子、增强子和终止信号。它们通常还包括核苷酸序列正确翻译所必需的序列。

术语“启动子”指通常位于编码区上游，且含有 RNA 聚合酶 II 结合位点，并启动 DNA 转录的非翻译 DNA 序列。启动子区还可包括作为基因表达调节物的其它元件。术语“最小启动子”指一种启动子元件，特别是一种失活的或在缺失上游激活元件下启动子活性急剧减少的 TATA 元件。

在本发明中术语“增强子”的含义可以是任何遗传元件如核苷酸序列，其增加了编码序列可操作地连接启动子的转录水平到一个高于当可操作地连接编码序列的启动子自身所实现的转录活化，即其增加了源自启动子的转录。

“嵌合腺病毒区”定义为含有一部分天然腺病毒基因组和一部分异源 DNA 序列的腺病毒载体的一部分。例如“嵌合腺病毒前导区”包含至少一部分的天然腺病毒前导区。例如 L1、L2、L3、L4、L5，还包含异源

DNA 序列。在一个实施例中，异源 DNA 序列包含转基因 CDS。在另一个实施例中，异源 DNA 序列包含转基因 CDS、分支点和剪接受体位点。

术语“转录调节元件”和“翻译调节元件”是影响核酸转录和/或翻译的那些元件。这些元件包括但不限于剪接供体和受体位点、翻译终止和起始密码子、以及腺苷酸化信号。

本文使用的“转录应答元件”或“转录调节元件”或“TRE”是一种多核苷酸序列（优选地是一种 DNA 序列），其包含一个或多个增强子和/或启动子和/或启动子元件，例如转录调节蛋白应答序列或增加可操作地连接的多核苷酸在容许 TRE 起作用的宿主细胞中的转录的序列。

“处于转录控制下”是本领域众所周知的术语，指多核苷酸序列（通常是 DNA 序列）的转录依赖于其被可操作地连接到有利于启动或促进转录的元件。

本文使用的术语“载体”指被设计用于在不同的宿主细胞之间进行转移的核酸构建体。载体可以是，例如，被设计用于分离、增殖和复制插入的核苷酸的“克隆载体”，被设计用于在宿主细胞中表达核苷酸序列的“表达载体”，或被设计用于产生重组病毒或病毒样颗粒的“病毒载体”，或包含多于一种类型载体的“穿梭载体”。任何用于基因导入的载体基本上都能用作“载体”，在其中具有期望序列的 DNA 已经被导入。质粒载体可用于本发明。当应用于本发明时，术语载体用来描述重组载体，如质粒或病毒载体（包括复制缺陷型或复制型病毒）。术语“载体”、“多核苷酸载体”、“多核苷酸载体构建体”、“核酸载体构建体”和“载体构建体”在此可交换地用来指如本领域技术人员所理解的任何用于基因转移的核酸构建体。

本文使用的与本发明病毒载体有关的术语“复制缺陷型”指无法进一步复制并包装其基因组的病毒载体。例如，当使用完整 E1 和 E4 编码区缺失或失活的腺病毒载体感染受试者细胞时，只有转基因在细胞中转录激活，才在患者细胞中表达异源转基因。然而，鉴于患者细胞缺少 Ad E1 和 E4 编码序列的事实，Ad 载体复制缺陷且在这些细胞中无法形成病毒颗粒。

术语“复制型”指能在特定细胞类型（“靶细胞”）如癌细胞中复制，并优先地影响那些细胞的细胞溶解的载体。已开发出能在癌细胞中选择性复

制并优先破坏那些细胞的特异性复制型病毒载体。对于多种细胞特异性复制型腺病毒构建体，其优先在某些细胞类型中复制（并因此破坏了这些细胞类型）。这样的病毒载体可称为“溶瘤病毒”或“溶瘤载体”，被认为能“溶细胞”或“致细胞病变”，并影响靶细胞的“选择性细胞溶解”。“复制型”或“溶瘤”病毒载体的例子描述于，例如 PCT 公开号 WO98/39466, WO95/19434, WO97/01358, WO98/39467, WO98/39465, WO01/72994, WO04/009790, WO00/15820, WO98/14593, WO00/46355, WO02/067861, WO98/39464, WO98/13508, WO20004/009790；美国临时申请序列号 60/511,812、60/423,203 和美国专利公开号 US 2001/0053352，其中每一篇都在此特别并入作为参考。

术语“复制条件性病毒”、“优先复制病毒”、“特异性复制病毒”和“选择性复制病毒”在此可交换使用，指在某些类型细胞或组织中优先复制但在其它细胞中复制水平较低或根本就不复制的复制型病毒载体和颗粒。在本发明的一个实施例中，病毒载体和/或颗粒在肿瘤细胞和/或诸如实体瘤或其它赘生物的异常增生组织中选择性复制。这样的病毒可称为“溶瘤病毒”或“溶瘤载体”，被认为能“溶细胞”或“致细胞病变”，并影响靶细胞的“选择性细胞溶解”。“优先复制”和“选择性复制”以及“特异性复制”可交换使用，并指与非靶细胞相比，在靶细胞中病毒复制更多。与非靶细胞相比，在靶细胞中病毒以更高的速率复制，例如至少约 3 倍高、至少约 10 倍高、至少约 50 倍高，以及在某些情况下，至少约 100 倍、400 倍、500 倍、1000 倍或甚至 1×10^6 倍高。在一个实施例中，病毒只在靶细胞中复制（也就是，在非靶细胞中根本就不复制或复制水平极低）。

本文使用的术语“质粒”指能在宿主细胞染色体外或作为宿主细胞染色体一部分能自发复制的 DNA 分子。本文的起始质粒是已经商用的，在不受限制的基础上公众能获得的，或可构建自依照在此所披露的和/或根据公开方法这样可获得的质粒。在某些情况下，如本领域技术人员所显而易见的，其它本领域已知的质粒能与在此描述的质粒可交换地使用。

术语“表达”指细胞中内源基因、转基因或编码区的转录和/或翻译。

“多腺苷酸化信号序列”是位于多腺苷酸化共有序列 AATAAA 之后的 RNA 转录物内源切割的识别区。多腺苷酸化信号序列提供了“polyA 位

点”，即通过后转录多腺苷酸化添加腺嘌呤残基到 RNA 转录物上的位点。一般来说，多腺苷酸化信号序列包括由侧接切割多腺苷酸化位点的两个识别元件组成的核心 poly (A) 信号（如 WO 02/067861 和 WO 02/068627 的图 1）。当完成与 poly (A) 位点强度有关的多腺苷酸化加工时，选择适合的多腺苷酸化信号序列应考虑多腺苷酸化信号序列的强度（Chao 等, *Molecular and Cellular Biology*, 1999, 19:5588-5600）。例如，强 SV40 晚期 poly (A) 位点比弱 SV40 早期 poly (A) 位点被切割得更快。如果需要的话，本领域技术人员应考虑选择更强的多腺苷酸化信号序列。原则上，对于本发明来说任何腺苷酸化信号序列都能使用。然而，在本发明的一些实施例中，终止信号序列是 SV40 晚期多腺苷酸化信号序列或 SV40 早期多腺苷酸化信号序列。一般而言，终止信号序列是分离自其基因源或通过合成构建并在适当位置插入本发明的载体中。

“多顺反子转录物”指含有多于一个蛋白编码区或顺反子的 mRNA 分子。包含两个编码区的 mRNA 称为“双顺反子转录物”。“5'-近侧”编码区或顺反子是其翻译起始密码子（通常是 AUG）接近多顺反子 mRNA 分子 5'-末端的编码区或顺反子。“5'-远侧”编码区或顺反子是其翻译起始密码子（通常是 AUG）不接近 mRNA 5' 末端起始密码子的编码区或顺反子。术语“5'-远侧”和“下游”同义地用来指不接近 mRNA 分子 5'末端的编码区。

本文使用的“内部核糖体进入位点”或“IRES”指一种元件，其促进指导内部核糖体进入顺反子（蛋白编码区）的起始密码子（例如 ATG），由此产生基因不依赖于加帽翻译（Jackson R J, Howell M T, Kaminski A (1990) *Trends Biochem Sci* 15 (12) :477-83) 和 Jackson R J 和 Kaminski, A. (1995) *RNA* 1 (10) :985-1000)。本发明包括使用任何一种能促进指导内部核糖体进入顺反子起始密码子的 IRES 元件。PCT 公开 WO 01/55369 描述了包括合成序列的 IRES 序列的例子，根据本发明这些序列亦可使用。本文使用的“处于 IRES 翻译控制下”指与 IRES 有关且以不依赖于加帽的方式进行的翻译。本领域已知的“IRES”的例子包括，但不限于，可获得自小 RNA 病毒的 IRES（Jackson 等, 1990, *Trends Biochem Sci* 15 (12) :477-483）；和可获得自病毒或细胞 mRNA 源的 IRES，例如免疫球蛋白重链结合蛋白（BiP）、血管内皮生长因子（VEGF）（Huez

等 (1998) *Mol. Cell. Biol.* 18 (11) :6178-6190)、成纤维细胞生长因子 2 和胰岛素类生长因子、翻译起始因子 eIF4G、酵母转录因子 TFIID 和 HAP4。在不同的病毒中还报道有 IRES, 例如心病毒、鼻病毒、口蹄疫病毒、HCV、弗里德小鼠白血病病毒 (FrMLV) 和莫洛尼鼠白血病病毒 (MoMLV)。本文使用的“IRES”包含 IRES 序列的功能性变体, 只要该变体能促进指导内部核糖体进入顺反子的起始密码子即可。在某些实施例中, IRES 是哺乳动物的。在其它实施例中, IRES 是病毒或原生动物的。在一个实施例中, IRES 可获得自脑心肌炎病毒 (ECMV) (购自 Novogen, Duke 等 (1992) *J. Virol* 66 (3) : 1602-1609)。在另一个在此披露的示范性实施例中, IRES 来自 VEGF。IRES 序列的例子描述于美国专利号 6,692,736。

“自加工切割位点”或“自加工切割序列”在此指一种 DNA 或氨基酸序列, 其中在翻译后, 包含自加工切割位点的多肽发生快速分子内 (顺式) 切割, 引起分离的成熟蛋白或多肽产物表达。这样的“自加工切割位点”也可称为后翻译或共翻译加工切割位点 (如 2A 位点)、序列或域。2A 位点、序列或域通过调节核糖体活性以促进酯键水解, 由此以容许进行分离的下游翻译产物合成的方式从翻译复合物上释放多肽, 从而显示对翻译的作用 (Donnelly, 2001)。可选择地, 2A 位点、序列或域通过顺式切割其自身 C-末端以产生主要切割产物, 从而显示“自我蛋白酶解”或“切割” (Furler; Palmenberg, *Ann. Rev. Microbiol.* 44:603-623 (1990))。

术语“剪接受体位点”或“3'剪接受体位点” (以下简称 3' SAS) 是一种可被改造进入野生型 Ad5 基因组不同位置的序列。通过利用腺病毒内源选择性剪接机制, 该位点的导入提供了一种便于表达一段异源 DNA 序列 (即转基因) 的方法。3' SAS 位点的一个例子是: TACTTATGAC TCGTACTATTGTTATTCATCCAG↓G (SEQ ID NO:45), 在此标记下划线的碱基是共有分支点 (其是可变的), A 指分支点, 其是高保守的, 下箭头指剪接位点 (使得箭头 (前) 的 5'序列被切除, 在箭头后的序列成为外显子的一部分)。

术语“分支点”和“分支点序列”可交换使用，指与剪接有关的核苷酸序列，并在本领域中被认为是套索形成位点的识别信号。当涉及 DNA 载体时，分支点序列是编码 RNA 分支点序列的 DNA 序列。

本文使用的“转基因”指一种在适当条件下在非天然环境或异源细胞中能通过重组技术可以表达的多核苷酸。在本发明中，转基因编码区被插入病毒载体中。在一个实施例中，病毒载体是一种腺病毒载体。转基因可来源于其要被表达、但从外源途径导入的相同类型的细胞，与相应的天然型相比是修饰的和/或由非天然位点表达的，或其可来源于异源细胞。“转基因”与“外源基因（exogenous gene）”、“异质基因（foreign gene）”、“异源编码序列”和“异源基因”是同义的。在上下文用于实施本发明的载体中，“异源多核苷酸”或“异源基因”或“转基因”是任何不存在于相应野生型载体或病毒中的多核苷酸或基因。转基因编码序列可以是天然存在的编码某一蛋白的序列。转基因编码序列还可以是非天然的编码序列。例如，本领域技术人员使用密码子选择图表，可容易地对编码序列再编码以优化在某些物种中用于表达的密码子。在一个实施例中，对于转基因，再编码序列依然与天然编码序列编码相同的氨基酸序列。在此提供了包含于本发明载体中的优选的转基因的例子。转基因可以是治疗基因。转基因并不是必需编码蛋白。

本文使用的“治疗”基因指当被表达时，赋予其中该基因被表达的细胞、组织或哺乳动物有益效果的转基因。有益效果的例子包括病症或疾病的病征或症状改善、病症或疾病预防或抑制、或赋予期望的特性。本领域已知为数众多的治疗基因例子，其中一些将进一步描述如下。

在上下文用于实施本发明的载体中，“异源”序列或元件是一种与相应的野生型载体或病毒无关或非源自相应的野生型载体或病毒的序列或元件。

在上下文用于实施本发明的载体中，“内源”序列或元件是天然于或源自相应的野生型载体或病毒的序列或元件。

“复制”和“增殖”可交换使用，指本发明病毒载体复制或增殖的能力。这些术语为本领域众所周知。对本发明来说，复制涉及病毒蛋白产生，且通常指病毒繁殖。可使用本领域标准的及在此描述的测定法测定复

制，例如病毒产量测定、爆发集落测定（burst assay）或色斑测定。“复制”和“增殖”包括任何直接或间接涉及病毒制造过程的活性，包括但不限于，病毒基因表达；病毒蛋白产生、核酸或其它成分复制；病毒成分包装为完整病毒和细胞溶解。

本文使用的“包装细胞”是一种能包装腺病毒基因组或修饰的基因组以产生病毒颗粒的细胞。其能提供缺失的基因产物或其等价物。因此，对于基因缺失的腺病毒基因组，包装细胞能提供补充功能，并能将腺病毒基因组包装为腺病毒颗粒。这种颗粒的产生需要复制基因组以及产生组装感染性病毒所必需的那些蛋白。这些颗粒还需要某些病毒颗粒成熟所必需的蛋白。这样的蛋白可由载体或包装细胞提供。

病毒载体的“生产细胞”是本领域众所周知的。生产细胞是一种在其中腺病毒载体被转移并复制及包装为病毒粒子的细胞。如果病毒载体必需基因缺失或失活，则生产细胞补充该失活的基因。腺病毒载体生产细胞的例子是 PerC.6 (Fallaux 等 Hum Gene Ther. 1998 年 9 月 1 日;9 (13) : 1909-17) 和 293 细胞 (Graham 等 J Gen Virol. 1977 年 7 月;36 (1) :59-74)。就选择性复制病毒来说，生产细胞可以是在其中病毒选择性复制的细胞类型。或则，此外生产细胞可表达在病毒载体中选择性受控或失活的基因。

术语“HeLa-S3”指可获得自美国典型培养物保藏中心（ATCC, Manassas, VA）并命名为 ATCC 号 CCL-2.2 的人宫颈肿瘤衍生的细胞系。HeLa-S3 是一种亲本 HeLa 系（ATCC CCL-2）的克隆衍生物。HeLa-S3 是在 1955 年被 T.T. Puck 等（J. Exp. Med. 103: 273-284 (1956)）克隆得到的。

“个体”是一种脊椎动物、哺乳动物或人。哺乳动物包括但不限于耕畜、体育动物、啮齿类、灵长类和宠物。“宿主细胞”包括能成为或已成为本发明的腺病毒载体受体的个体细胞或细胞培养物。宿主细胞包括单个宿主细胞的后代，且由于天然的、偶然的或有意的突变和/或改变，该后代无需完全相同（在形态学或总 DNA 上）于原始的亲本细胞。宿主细胞包括使用本发明的腺病毒载体体内或体外转染或感染的细胞。

本文使用的“细胞毒性”是本领域众所周知的术语，指其中细胞通常的生物化学或生物学活性受损（即被抑制）的状态。这些活性包括但不限于

新陈代谢；细胞复制；DNA 复制；转录；翻译；分子摄取。“细胞毒性”包括细胞死亡和/或细胞溶解。指示细胞毒性的测定法是本领域已知的，例如染料排除、³H-胸苷摄取和噬菌斑测定。

本文使用的术语“赘生性细胞”、“瘤形成”、“肿瘤”、“肿瘤细胞”、“癌（carcinoma）”、“癌细胞（carcinoma cells）”、“癌（cancer）”和“癌细胞（cancer cells）”（可交换使用）指显示相对自发生长的细胞，使得它们显示异常的生长表型，具有细胞增殖明显失控的特征。赘生性细胞可以是恶性或良性的。

腺病毒载体

本文使用的术语“腺病毒”和“腺病毒颗粒”包括任何及所有可被分类为腺病毒的病毒，包括任何感染人或动物的腺病毒，包括所有已知的及之后发现的组、亚组和血清型。因此，除非另有说明，本文使用的“腺病毒”（“Ad”）和“腺病毒颗粒”指病毒自身或其衍生物，并覆盖了所有的血清型和亚型，并且是天然存在的或重组形式。这样的腺病毒可以是野生型的或可以是以多种本领域已知的或在此披露的方法所修饰的。这样的修饰包括改变包装在颗粒中的腺病毒基因组以便制备感染性病毒。这样的修饰还包括本领域已知的缺失，例如缺失一个或多个腺病毒复制必需基因，如 E1a、E1b、E2a、E2b、E3 或 E4 编码区。术语“复制必需基因”指对于在靶细胞中复制病毒载体，需要其转录的核酸序列。例如，在本发明的腺病毒载体中，复制必需基因可选自 E1a、E1b、E2a、E2b 和 E4 基因。该术语还包括复制特异性腺病毒；即在某些类型的细胞或组织中优先复制，但在其它类型的细胞或组织中复制水平较低或根本就不复制的病毒。有时，这样的病毒称为“溶细胞性”或“致细胞病变”病毒（或载体），如果它们对赘生性细胞具有这样的作用，则称为“溶瘤”病毒（或载体）。

本发明示范性的腺病毒载体包括，但不限于 DNA、包在腺病毒外壳中的 DNA、包装在另一种病毒或病毒样形式（例如单纯疱疹和 AAV）中的腺病毒 DNA、包在脂质体中的腺病毒 DNA、与多聚赖氨酸复合的腺病毒 DNA、与合成的聚阳离子分子复合的腺病毒 DNA、与铁传递蛋白缀合

的腺病毒 DNA、或与诸如 PEG 的化合物复合以在免疫学上“掩蔽”抗原性和/或增加半衰期的腺病毒 DNA、或与非病毒蛋白缀合的腺病毒 DNA。

在上下文腺病毒载体中，术语“5'”与“上游”可交换使用，并指在反向末端重复（ITR）的左侧方向。在上下文腺病毒载体中，术语“3'”与“下游”可交换使用，并指在 ITR 的右侧方向。

用于产生表达插入序列的腺病毒载体的标准系统是本领域已知的，并且商业可购，例如获得自得 Clontech（Clontechniques（2000年1月）p. 10-12）的 Adeno-X 表达系统。

本发明的腺病毒载体包括复制缺陷型和复制型载体。在靶细胞中，复制缺陷型载体不复制或复制水平十分低。在一个实施例中，复制缺陷型载体一般通过缺失或突变部分或全部的编码区，在 E1a、E1b、E2a、E2b 或 E4 中具有至少一个失活的编码区。增殖这些载体的方法是本领域众所周知的。

本发明涉及使用所有的腺病毒血清型以构建腺病毒载体和根据本发明的病毒颗粒。根据本发明可被使用的腺病毒株包括所有的腺病毒血清型。现在，腺病毒血清型 1-47 可获得自美国典型培养物保藏中心（ATCC, Manassas, VA），且本发明包括所有其它任何来源所能获得的腺病毒血清型。根据本发明可被使用的腺病毒可以是人源或非人源的。举例来说，腺病毒可以是亚组 A（如血清型 12、18、31）、亚组 B（如血清型 3、7、11、14、16、21、34、35）、亚组 C（如血清型 1、2、5、6）、亚组 D（如血清型 8、9、10、13、15、17、19、20、22-30、32、33、36-39、42-47）、亚组 E（血清型 4）、亚组 F（血清型 40、41）、或任何其它腺病毒血清型。说明书全文中所提到的是 5 型腺病毒中的特定核苷酸。基于本领域技术人员所能获得的常规知识，本领域技术人员可以确定其它血清型中的相应核苷酸，并因此构建腺病毒血清型不同于血清型 5 的类似的腺病毒载体。在一个优选的实施例中，腺病毒核酸主链源自腺病毒血清型 2（Ad2）、5（Ad5）或 35（Ad35），或者嵌合腺病毒主链包含腺病毒血清型 2（Ad2）或 5（Ad5）部分与腺病毒血清型 35（Ad35）部分的组合。

腺病毒血清型 2 和 5 的 DNA 和蛋白序列可见于基因库 (GenBank) 保藏登记号 NC_001405 (Ad2) 和 AY339865 (Ad5), 两者全文在此并入。随同完整的基因组 DNA 序列一起, GenBank 条目包括了有用的详细资料, 例如参考文献、剪接信号位置、多腺苷酸化位点、TATA 信号、内含子、每一标识基因的起始和终止密码子、蛋白序列、每一基因的 cDNA 以及在存在于全文文献中的一列序列变体。同样地, 对于本发明特别感兴趣的是, 可从所指示的每一基因或区的剪接位点和多腺苷酸化切割位点推断出每一区的 mRNA 结构, 以及在这些 GenBank 记录中相关出版物的引用列表。

以下参考文献提供了腺病毒基因位置、基因剪接、转录元件 (如剪接位点、多腺苷酸化信号序列) 位置的详细资料: Nevins 和 Chen-Kiang, *Adv Virus Res.* 1981;26:1-35; Akusjarvi 和 Stevenin, *Curr Topics Microbiol Immunology* 272:253-86 (2003); Prescott 和 Falck-Pedersen, *Mol Cell Biol* 1994 14:4682-4693; Muhlemann 等 1995 *J Virol* 69:7324-7327; Nevins 和 Wilson, *Nature.* 1981 年 3 月 12 日;290 (5802) :113-8; Prescott 和 Falck-Pederson, *J Biol Chem* 267:8175; 1992; Larsson, Svensson 和 Akusjarvi *J Mol Biol* 225:287; 1992; Farley, Brown 和 Leppard, *J Virol* 78:1782; 2004。对于其基因组两条链 (上链及下链) 上的基因, 腺病毒编码基因和加工信号。

当进行任何变化时, 注意在互补链上应不存在不希望有的改变。同样地, 有时转基因相对较小, 则因此多个转基因有可能被插入到一个腺病毒载体基因组中。只要两条 DNA 链上必需基因和加工信号都保持完整, 转基因就能基本上插入 Ad 基因间的任何位置。可选择地, 在某些情况下, 期望改变现有基因和/或信号以改变病毒基因表达来适应载体的应用。同样地, 对于新的 CDS, 可能需要添加分支点和剪接受体位点或其它加工信号, 与转基因一起用于新导入 mRNA 的正确加工。

在一个优选的方面, 腺病毒载体是复制型或复制条件型的。这样的载体能在靶细胞中复制。复制型病毒包括野生型病毒和被改造以在靶细胞中复制的病毒。这些包括与其它细胞相比用来在一种类型的靶细胞中特异性或优先复制的载体。靶细胞可以是某种细胞型、组织型或具有某种细胞

状态。其中复制取决于细胞型、组织型或细胞状态的复制型腺病毒载体进一步描述如下。

在一个实施例中，本发明的腺病毒载体包含处于赋予腺病毒选择性复制的异源转录调节元件（TRE）转录控制下的一个或多个腺病毒复制必需基因。所述腺病毒复制必需基因是一种早期基因，选自 E1A、E1B、E2a、E2b 和 E4。在另一个实施例中，一个或多个额外的 TREs 可操作地连接一个或多个腺病毒复制必需基因或转基因，例如治疗基因。

腺病毒 E1B 19-kDa 区指编码 E1B 19-kDa 产物的腺病毒 E1B 基因的基因组区。根据野生型 Ad5，E1B 19-kDa 区是位于核苷酸（nt）1714 和 nt 2244 之间的 261bp 区。E1B 19-kDa 区已描述于例如，Rao 等, *Proc. Natl. Acad. Sci. USA*, 89:7742- 7746。在一个实施例中，本发明包括 E1B 19-kDa 区部分或全部缺失，以及其中 E1B 19-kDa 区突变的实施例，只要缺失或突变减少或消除与 E1B-19kDa 有关的细胞凋亡即可。

在本发明的一个实施例中，腺病毒载体优先在癌靶细胞中复制，通过比较癌靶细胞与非靶细胞、非癌细胞、正常或对照细胞中的复制水平（如细胞溶解或细胞杀死和/或滴度）以表明其复制优先。靶细胞中特定腺病毒滴度和非靶细胞（或“TRE 失活的”细胞）中腺病毒滴度的比较表明复制优先在靶细胞中增加了和/或在非靶细胞中降低了。

腺病毒载体还可包括一个或多个异源 TREs，其可以或不可以可操作地连接相同的基因。例如，细胞型特异性、细胞状态特异性或组织型特异性 TRE（所有在此描述的都称为“细胞特异性”或“靶细胞特异性”型 TREs）可与相同或不同类型的第二 TRE 并置。“并置”指第一 TRE 和第二 TRE 转录控制相同的基因。对于这些实施例，多于一个的 TRE 可以任何一种构型存在，包括但不限于：（a）彼此邻接（即毗连）；（b）都位于转录受控基因的 5'（其中在两个 TREs 之间可具有间插序列）；（c）一个 TRE 可位于基因的 5' 而另一个 TRE 位于该基因的 3'。

可使用本领域标准的重组技术制备本发明的病毒载体。修饰复制型或非复制型病毒载体的方法是本领域众所周知的，并描述于此及在此引用的出版物。描述于此的用于将转基因和期望的转录元件克隆入腺病毒的多种方法是标准方法，且为本领域众所周知。如在此描述的，转基因和期望

的转录元件被克隆入腺病毒载体基因组中的不同位点。例如，在本领域中存在多种含有腺病毒基因组不同部分的质粒，包括含有完整腺病毒基因组的质粒。这些质粒的构建在本领域中也得到了充分的描述（如 US20030104625）。一旦用于转基因插入的位点被选择，那么适当质粒就可被用来进行修饰。然后，该修饰可通过例如同源重组或体外连接导入全长腺病毒载体基因组。同源重组可在哺乳动物细胞（如 PerC6）或细菌细胞（如大肠杆菌，参见 WO9617070）中发生。病毒载体基因组操作能可选地或另外包括众所周知的分子生物学方法，包括但不限于，聚合酶链反应（PCR）、PCR-重叠延伸拼接（PCR-SOEing）、限制酶切消化。如果使用同源重组，则两个质粒通常具有至少约 500 bp 的序列重叠，尽管较小的重叠区会重组，但通常效率较低。视所需，每个质粒可独立地操作，接着共转染入感受态宿主中，提供适合于腺病毒载体增殖的补充基因。使用诸如正离子脂质体或磷酸钙的适当转导方法，通常可将质粒导入适合的宿主细胞（如 293、PerC.6、Hela-S3 细胞）中。或者，还可进行腺病毒基因组右侧和左侧部分的体外连接以构建含有所有腺病毒基因组复制必需部分的重组腺病毒衍生物。Berkner 等（1983）*Nucleic Acid Research* 11: 6003-6020；Bridge 等（1989）*J. Virol.* 63: 631-638。

为了方便起见，可使用提供了腺病毒必需部分的质粒。质粒 pXC.1（McKinnon（1982）*Gene* 19:33-42）含有野生型 Ad5 的左侧末端。pBHG10（Bett 等（1994）；Microbix Biosystems Inc., Toronto）提供了具有 E3 缺失的 Ad5 的右侧末端。在病毒载体中 E3 缺失提供了更多插入异源序列的空间。E3 基因位于 E4（右链）的相反链上。pBHG11 提供了甚至更大的 E3 缺失，缺失了额外的 0.3 kb（Bett 等（1994））。可选择地，使用提供了具有全长 E3 的 Ad5 右侧末端的 pBHGE3（Microbix Biosystems, Inc.）。

将多核苷酸包装入腺病毒颗粒中的方法是本领域已知的，也描述于 PCT/US98/04080。优选的包装细胞是已被用来限制能产生野生型腺病毒颗粒的同源重组的那些细胞。能用来产生本发明的腺病毒颗粒的细胞包括人胚胎肾细胞系 293（Graham 等, *J Gen. Virol.* 36:59-72（1977））、人胚胎成视网膜细胞系 PER.C6（美国专利号 5,994,128 和 6,033,908；Fallaux 等,

Hum. Gene Ther. 9: 1909-1917 (1998)) 以及人宫颈肿瘤衍生的细胞系 HeLa-S3 (PCT 申请号 US 04/11855)。病毒载体可以多种方式转移至靶细胞, 包括但不限于脂质体、本领域众所周知的常用转染方法(例如磷酸钙沉淀或电穿孔)、直接注射和静脉输注。在很大程度上, 转移方式取决于特定的载体(包括其类型)以及靶细胞的类型和位置(即细胞是否在体外或体内)。

本发明还提供了包含根据本发明的重组病毒载体的重组腺病毒颗粒。在一个实施例中, 腺病毒颗粒的衣壳蛋白包含靶向配体。在一种方法中, 衣壳蛋白是尾丝蛋白或 pIX。在一方面, 衣壳蛋白是尾丝蛋白, 且配体在尾丝蛋白的 C 末端(羧基末端)或 HI 环中。腺病毒载体颗粒还可包括其它对尾丝蛋白的突变。在一个另外的实施例中, 通过使用来自另一种腺病毒血清型的尾丝结(fiber knob)取代尾丝结以靶向病毒。这些突变的例子包括, 但不限于描述于美国申请号 10/403,337; 美国申请公开号 2004/0002060; PCT 公开号 WO 98/07877; WO 99/39734; WO 00/67576; WO 01/92299; 和美国专利号 5,543,328; 5,731,190; 5,756,086; 5,770,442; 5,846,782; 5,962,311; 5,922,315; 6,057,155; 6,127,525; 6,153,435; 6,455,314; 6,555,368 和 6,683,170 以及 Wu 等(J Virol. 2003 年 7 月 1 日;77 (13):7225-7235)中的那些。这些包括, 但不限于减少病毒载体颗粒与特定细胞类型或多于一种的细胞类型结合的突变、增强病毒载体颗粒与特定细胞类型或多于一种细胞类型结合的突变、和/或降低哺乳动物中对腺病毒载体颗粒免疫应答的突变。

在 C 组腺病毒中, 腺病毒主要晚期转录单位(MLTU)分为五个区。每个区特点在于都具有自己的多腺苷酸化信号序列, 且通常由几个不同的剪接信使组成。所有的晚期信使都在其三联前导序列(TPL)的 5'末端被剪接。每个区的描述如下。

在本发明示范性的实施例中, 转基因插入在至少一个如下位置中:
A) L1 区, 插入 52、55K 和 IIIa CDS 之间, 或 pIIIa CDS 之后和 L1 区多腺苷酸化信号序列上游之间(图 3); B) L2 区: 插入 pV CDS 和 Mu CDS 之间, 或 pVII CDS 和 pV CDS 之间, 和/或五邻体之后和 pVII 之前(图 4); C) L3 区: 插入 pVI 上游, 或 pVI 和六邻体之间, 或六邻体和

23K3 之间，或 23K 基因之后和多腺苷酸化信号序列上游之间（图 5-7）；D）L4 区：使用如其它区所述的类似策略，可能插入 L4 区多腺苷酸化信号序列上游。然而，在 L4 和 E2 转录单位和加工信号之间存在着重要的重叠。因此，本领域技术人员会以不显著破坏其它转录单位和加工信号的方式插入转基因，当然除非期望所述破坏时。E）L5 区：在 Ad5 L5 区中只有一个尾丝基因编码的蛋白。终止密码子嵌在多腺苷酸化信号序列中。转基因插入在 L5 尾丝 CDS 的上游，或如果插入下游，则对于尾丝 CDS，多腺苷酸化信号序列将被重建/添加，突变天然多腺苷酸化信号序列以保持终止密码子并失活多腺苷酸化功能。F）E3 区：转基因插入在没有缺失的 E3 区。例如，转基因插入位点包括 E3 基因上游或下游的位点。G）其它存在于 Ad 基因组中的转录单位，例如编码 pIX 和 IVa2 的中间区，可以类似的方式得以使用。H）可添加到基因组中的新转录单位。例如，可添加新的晚期区到晚期转录单位，形成“L6 区”。

在一个实施例中，转基因作为 L1、L2、L3、L4 或 L5 转录物的一部分表达。

在一个实施例中，载体中的转基因利用了（即可操作地连接）为至少一个病毒 mRNAs 所使用的相同的 polyA 信号。例如，当在主要晚期转录单位中插入转基因时，插入转基因的方法并不创建新的或第六前导区序列。相同的 polyA 信号可以是天然 Ad polyA 或异源 polyA 信号。

对多腺苷酸化信号序列附近序列的修饰可影响切割效率。因此，对这些区的修饰可增加、减少或产生相当的切割效率。当选择插入位点时，本领域技术人员能考虑到期望的表达时间和转基因的表达量。

在本发明的一个实施例中，IRES 或自加工切割序列可操作地连接插入多腺苷酸化信号序列上游和腺病毒编码序列下游的转基因。多腺苷酸化信号序列通常是腺病毒多腺苷酸化信号序列，但可以是异源多腺苷酸化信号序列。在一个实施例中，所述异源多腺苷酸化信号序列替代了腺病毒多腺苷酸化信号序列的功能。在本发明的一个实施例中，IRES 或自加工切割序列和转基因插入在编码多于一种蛋白的腺病毒转录单位或前导区序列（如 L1、L2、L3、L4）中（如由于差别剪接）。该实施例优点在于 IRES 或自加工切割序列对来自所述转录物或前导区的每一 mRNA 都起作

用，换言之，除腺病毒编码序列（CDS）之外，来自所述转录单位的每一 mRNA 应可操作地编码并表达转基因。在一个例子中，其后连接有转基因的 IRES 或自加工切割序列插入 L2 多腺苷酸化信号的上游，L2 区产生的所有四种类型的 mRNAs 应都含有所插入的序列，因此所有四种类型的 mRNAs 都应表达转基因。不受理论束缚，人们相信 IRES 或自加工切割序列和转基因与多腺苷酸化信号越接近，转基因的表达水平就越高。当选择插入位点时，本领域技术人员能考虑到期望的表达时间和转基因的表达量。例如，如果期望在相对于病毒复制周期的中间时间表达转基因，则 IRES 或自加工切割序列转基因盒也可插入 pIX 或 IVa2 的 polyA 信号上游。能获得各种类型的 IRESs 和自加工切割序列，各自介导翻译的效率不同。本领域技术人员能选择具有已知效率的特定 IRES 或自加工切割序列用于调节转基因表达效率。当添加 IRES 或自加工切割序列和转基因时，务必小心避免破坏其它腺病毒序列，包括编码其它链（即反义链）的那些，除非期望破坏某些腺病毒基因。鉴于对腺病毒基因表达的大量描述，本领域技术人员能确定会或不会破坏腺病毒基因表达的位点。

L1 区

如图 1-3 所示，L1 区包括 52/55K 和 IIIa 编码区。在剪接受体位点之后的 ATG 上游，两个基因都只具有一个碱基。在所有的晚期区中，每个区所编码的 mRNAs 在其剪接受体位点与三联前导序列相连。已鉴定了在 52/55K 和 IIIa 编码区之间剪接增强子序列（25 个碱基对序列）的存在和位置。此外，剪接阻遏区正好位于剪接增强子上游。L1 区上游是 VA RNA 编码区和位于互补链上的 E2B 编码区。事实上，VA RNA II 结束于自 52/55K mRNA 剪接受体位点的约 10 个碱基对。L2 区下游的剪接受体位点位于 L1 多腺苷酸化信号序列上游。这些重叠部分应被考虑，且当转基因插入在 L1 区时，通常要完整保留。尽管只产生 52/55K 基因的可检测水平，但在感染早期 L1 mRNA 就开始合成了。在进入晚期后，IIIa 编码区被翻译，且 52/55K 翻译水平更低。此外，这种调节改变的机制应归于受进入感染晚期影响的差别剪接位点的使用。

转基因能潜在地位于两个 L1 编码区之间、或 52/55K 编码区上游、或紧接 IIIa 编码区下游以及 L1 多腺苷酸化信号序列上游。在一个实施例

中，在转基因和/或腺病毒编码区下游添加分支点序列和剪接受体位点。在另一个实施例中，自加工切割位点添加在 IIIa 区下游并可操作地连接 IIIa 区，之后自加工切割位点可操作地连接位于自加工切割位点下游的转基因。同样地，当设计 IIIa 编码区下游插入时，务必小心避免破坏 L1 多腺苷酸化信号序列和 L2 剪接受体位点。或者，视所需可恢复或改变这些信号以改变蛋白表达。考虑到类似于病毒蛋白，插入 52/55K 上游和下游的转基因会在早期表达，而插入 IIIa 下游的转基因则会在感染晚期表达。取决于例如剪接增强子和阻遏区的布置，其可得到改变。

在另一个实施例中，将可操作地连接转基因的 IRES 插入 IIIa 编码区下游和 L1 多腺苷酸化信号序列或异源多腺苷酸化信号序列上游（即可操作地连接）。

L2 区

如图 1-2 和 4 所示，L2 区编码四个已鉴定的编码区：五邻体（aka III）、proVII（aka pVII）、pV 和 pX（有时称为 pV 前体或“mu”）。五邻体具有在其起始密码子上游 2 个碱基处的剪接受体位点。proVII mRNA 的剪接位点嵌在五邻体编码区中，其起始密码子在五邻体终止密码子下游 7 个碱基对处。proVII 的终止密码子、pV mRNA 的剪接位点和 pV 的起始密码子分别间隔 46 和 23 个碱基对。pX 及其 ATG 起始密码子在 pV 终止密码子下游 124 个碱基处。L2 区的多腺苷酸化信号序列在 pX 终止密码子后 26 个碱基处，在 polyA 信号后约 12 个碱基处发生切割。

转基因插入的选择包括但不限于：a) 插入 pV 或 pX CDSs 后，和 L2 多腺苷酸化信号序列上游，b) 插入 pV mRNA 剪接受体位点下游和 pV CDS 上游，和 c) 插入五邻体终止密码子和 pVII 起始密码子之间。当转基因插入方法和表达依赖于选择性剪接时，期望在每个编码区都具有剪接受体位点，可以是现有的或外源添加的序列。所有的 L2 mRNA 信使通常都使用 L2 多腺苷酸化信号，但是在某些实施例中，可利用非天然 polyA 信号。

使用自加工切割位点，转基因还可表达自 L2 区的这些区。在一个实施例中，自加工切割位点可操作地连接 Ad CDS（如五邻体、proVII、pV

和 pX) 的 3'末端, 转基因 CDS 可操作地连接自加工切割位点的 3'末端。在一个可选的实施例中, 转基因插入在 Ad CDS 上游, 其中自加工切割位点可操作地连接转基因的 3'末端, Ad CDS 可操作地连接自加工切割位点的 3'末端。在该情况下, 转基因与适当的转录元件(如在此描述的)一起插入使得其从载体转录。这可包括插入可操作地连接 Ad 剪接位点和分支点的转基因。在一个实施例中, Ad 剪接位点和分支点是在天然病毒中连接 Ad CDS 下游的位点。在另一个实施例中, 转基因可操作地连接其中之一或两者为异源的剪接位点和分支点。

在另一个实施例中, 可操作地连接转基因的 IRES 插入 pV 编码区下游和 L2 polyA 信号或异源多腺苷酸化信号序列上游(即可操作地连接)。

L3 区

如图 5-7 所示, L3 区含有 pVI、六邻体(II)和 23K(病毒蛋白酶)编码区。pVI mRNA 的剪接受体位点位于 L2 多腺苷酸化位点下游和 pVI 起始密码子上游一个碱基处。pVI 终止密码子位于六邻体信使剪接受体位点上游约 50 个碱基处, 六邻体起始密码子位于自其 mRNA 剪接受体位点的约 35 个碱基处。23K mRNA 的剪接受体位点包含在六邻体编码序列中, 位于六邻体终止密码子上游约 95 个碱基处。23K 起始密码子位于六邻体终止密码子下游的 32 个碱基处。23K 终止密码子位于 L3 polyA 信号上游约 25 个碱基处。L3 多腺苷酸化区与 E2A 多腺苷酸化信号的重叠位于互补链上。在大多数情况下, 期望在两条链上保留完整的多腺苷酸化信号序列用于有效加工。

转基因插入在 L3 区中存在多种选择: a) pVI 编码区上游, b) pVI 和六邻体编码区之间, c) 六邻体和 23K 编码区之间和 d) 23K CDS 末端。注意到 L3 polyA 信号十分强且六邻体 mRNA 大量表达。预计插入 L3 中的转基因还会大量表达。同样地, 对于在六邻体和 23K 编码区之间的插入, 对于转基因 mRNA 期望使用六邻体编码区中的剪接受体位点(通常用于 23K mRNA), 对于 23K 信使合成则期望添加新的分支点和剪接受体位点。在另一个实施例中, 转基因与具有自加工切割位点的六邻体

CDS 连接。在另一个实施例中，可操作地连接转基因的 IRES 插入 23K 编码区下游和 L3 polyA 信号或异源多腺苷酸化信号序列上游（即可操作地连接）。

L4 区

如图 1 和 2 所示，L4 区含有 100K、33K 和 pVIII 蛋白编码区。33K CDS 的一部分与 100K CDS 重叠。尽管在这两个基因之间存在非编码碱基对，但并非插入转基因的期望位置，因为这可能破坏 33K 内含子或改变其使用。pVIII CDS 和 polyA 区与 E2 和 E3 启动子重叠。在互补链上的是 E2A 和 E2B 转录单位的前导区序列。因此，当利用该区进行转基因插入时，设计者应当意识到这一点。有可能在该区的任意插入或破坏能改变或破坏 E2、E3 和/或 L4 区中必需基因的表达。

在另一个实施例中，可操作地连接转基因的 IRES 或自加工切割位点插入 100K 编码区下游和 L4 polyA 信号或异源多腺苷酸化信号序列上游（即可操作地连接）。在该情况下，因为与 L4 polyA 序列重叠，E3 12.5K CDS 可能会被中断。然而，某些该基因缺失的突变体在体外并不显示有表型改变，因此该转基因插入位置/方法能产生功能性腺病毒载体。

L5 区

如图 1 和 2 所示，Ad5 中 L5 区仅编码尾丝蛋白。尾丝 mRNA 的剪接受体位点位于尾丝起始密码子上游 2 个核苷酸（nts）处。尾丝终止密码子嵌于 L5 polyA 信号中。转基因可插入尾丝编码区上游或下游。然而，对于下游插入，天然 L5 多腺苷酸化信号序列被修饰，使得该序列不再提供多腺苷酸化功能。在一个实施例中，在 mRNA 合成中插入剪接受体位点和转基因。然后，通过在插入的转基因下游，多腺苷酸化信号得以恢复供改变的 L5 区使用。在另一个实施例中，自加工切割位点可操作地连接尾丝 CDS 下游末端，转基因 CDS 插入自加工切割位点下游并可操作地与其连接。

额外的晚期区

另一种将转基因插入在诸如 Ad 的病毒载体中的方法是插入额外的晚期区。可通过例如恰好插入现有区或转录单位（例如 L3），由分支点和剪接受体位点、转基因和多腺苷酸化信号（以及任何其它必需信号）组成的盒的下游而得以进行。该盒还可由 IRES 或可操作地连接第二转基因的自加工切割位点组成。第二转基因可编码与第一转基因相同或不同的蛋白。就相同的蛋白来说，优选转基因的编码序列“再编码”。换言之，使用不同的密码子来编码相同的氨基酸。能在 DNA 水平上减少两个转基因之间的同源性数值，因此减少或消除两个转基因之间的同源重组。

E3 区

腺病毒中 E3 区编码已鉴定的蛋白 12.5K、6.7K、gp19K、ADP、RID- α 、RJD- β 和 14.7K 的基因。在一个实施例中，转基因插入在 gp19K 编码区下游和 ADP CDS 上游。有趣的是，在 Ad5 病毒中有 128 bps 位于 gp19K 和 ADP 编码区之间（保藏登记号 AY339865 的 Ad 基因组中碱基对 29209-29336；SEQ ID NO:41）。

在本发明的一个实施例中，这些 128 bps 或这些 128 bps 的大多数都从 Ad5 载体上缺失了。在一个实施例中，Ad5 载体含有 gp19K 和/或 ADP 编码区，且缺失这些 128 bps。Ad2 病毒在该区并不含有这些 128 bps 或类似序列。在另一个实施例中，转基因插入在 14.7K 编码区后和 E3B 多腺苷酸化信号序列上游。在另一个实施例中，转基因插入在 ADP 编码区后。然而，ADP 终止密码子却嵌于 E3A polyA 位点中。因此，应修饰 polyA 位点以使得该序列不再提供多腺苷酸化功能，但在 ADP 编码序列读框中仍保留终止密码子。转基因可插入（如与剪接受体位点或自加工切割位点一起）ADP CDS 下游。然后，多腺苷酸化信号序列插入转基因编码区下游。所插入的多腺苷酸化信号序列可以是天然的 E3A polyA 位点或额外的 polyA 位点（如 SV40 polyA 位点）。

E3 编码区并不是病毒复制所必需的。因此，在腺病毒载体中 E3 编码区不是必需的。结果，本发明的病毒载体可具有一个或多个 E3 编码区缺失，或可选择地，可包含所有的 E3 编码区。在本发明的实施例中，E3 14.7、14.5 或 10.4k 编码区或其组合保留在腺病毒载体中。在本发明的一

个实施例中，不插入取代 E3 CDS 的转基因。例如，转基因插入在两个天然 E3 CDSs 之间，其中在天然病毒中所述两个天然编码序列彼此邻接，且没有其它的 E3 CDS 位于它们之间。在某些实施例中，本发明的腺病毒载体含有可操作地连接转基因的异源剪接受体位点，其作为 E3 转录单位的一部分转录。例如，异源剪接受体位点可操作地连接转基因，并且都插入 E3 区中。

在另一个实施例中，插入可操作地连接转基因 CDS 的自加工切割位点，使得该自加工切割位点还可操作地连接腺病毒 E3 CDS。转基因 CDS 可位于自加工切割位点下游，E3 CDS 位于自加工切割位点上游。在另一个实施例中，E3 CDS 可位于自加工切割位点下游，转基因 CDS 位于自加工切割位点上游。

在另一个实施例中，可操作地连接转基因的 IRES 插入在 14.7K 编码区下游和 E3B polyA 信号或异源多腺苷酸化信号序列上游（即可操作地连接）。在一个实施例中，可操作地连接到转基因的 IRES 插入在 ADP 编码区下游和 E3A polyA 信号或异源多腺苷酸化信号序列的上游（即可操作地连接）。E3 区编码不为病毒复制所需的蛋白，因此对于某些类型的腺病毒载体及其应用不是必要的。因此，在另一个实施例中，一个或多个 E3 编码序列缺失或突变，由可操作地连接转基因的 IRES 组成的异源 DNA 分别被插入在 E3A 或 E3B 多腺苷酸化信号序列和紧接着 E3 A 或 E3B 多腺苷酸化信号序列 5' 的 E3 编码序列之间。

任何一种在此描述的转基因插入方法都可与任何另外一种在此或其它地方所描述的转基因插入方法组合使用，只要插入位点一致即可。例如，可操作地连接剪接受体位点和分支点的第一转基因可插入在转录单位中。然后，所插入的第一转基因能可操作地连接插入在所述第一转基因下游的自加工切割位点，接着第二转基因插入在自加工切割位点下游并与其可操作地连接。本领域技术人员应当理解的是，能依赖于本文教导，使用所披露的转基因插入和表达方法的组合插入转基因，所有方法和插入位点都包含在本发明中。

如果要将腺病毒载体 DNA 包装到病毒粒子中，那么必须小心不要超过病毒的包装容量。例如对于 Ad5，当基因组 DNA 大于野生型 Ad5 基因

组的尺寸约 105%时，包装效率就大大降低（Bett 等 *Virol.* 1993 年 10 月;67（10）:5911-21）。不顾稳定的更大的 Ad5 载体已被报道。

为简单起见，除非另作说明，在说明书上下文中指向 Ad 血清型 5。本领域技术人员无需过度试验就能容易地推断出在其它腺病毒或其它病毒中用于转基因的等价位点

取决于转基因的特性，可期望在感染早期或晚期表达；此外，亦可期望具有高或低的表达水平。例如，在某些情况下，优选在病毒 DNA 复制后的感染晚期表达。如果载体是具有驱动早期基因表达的肿瘤特异性启动子的溶瘤 Ad 载体，则晚期表达提供了一种额外的调节机制，因为晚期基因表达需要 DNA 复制。如果转基因插入在晚期转录物中，则表达是特异性，因为对于靶细胞病毒复制是特异性的，如果 Ad 载体被设计为复制选择性的话。

另一种可用于帮助调节转基因表达或改变病毒基因表达的机制是修饰与共有序列相比的所添加元件（如分支点序列和/或剪接受体位点或自加工切割位点）的序列。其改变了使用效率。例如，就剪接受体位点来说，如果期望高表达水平，则应使用与共有剪接受体序列十分接近的序列。共有剪接受体位点序列确定为(T/C)₈, N, C/T, A,G,G（Mount, *Nucleic Acids Res* 10:459; 1982）。已进行了改变切割效率的突变分析（Rosigno 等 *J Biol Chem* 268: 11222; 1993; Lee 等 *Gene Therapy* 11 : 94; 2004）。应当指出在病毒进入晚期后，通常所接受的共有序列定义和使用可以并进行改变（Akusjarvi 和 Svevenin, *Curr Top Microbiol Immunol* 272: 253; 2003; Nevins 和 Wilson, *Nature* 290:113; 1981; Akusjarvi 和 Persson, *J Virol* 38: 469; 1981）。例如，在感染晚期非共有剪接受体位点的使用得以增加（Muhlemann 等 *J Virol* 69: 7324; 1995）。大多数的这些改变是由于病毒蛋白对用于转录和翻译的细胞结构的作用之故。

除剪接受体位点之外，分支点序列亦可包含在本发明的腺病毒载体中。分支点序列是内含子有效剪接所必需的（Vandenbroucke 等, *BMC Genomics* 3: 13; 2002; Hall 等, *PNAS* 85:704-708; 1988; Harris 等 *Nucl Acid Res* 18（10） 3015-3019 1990）。分支点与剪接受体位点的距离也影响剪接效率，且通常为，但并不总是剪接受体位点上游 10-50 个碱基对，

更经常地是 18-37 个碱基对 (Vandenbroucke 等, 2002; Hall 等 1988; Harris 等 1990)。据信分支点序列的共有序列是 YNYURAY (SEQ ID NO:40) (在此 Y 是嘧啶, N 是任何核苷酸以及 R 是嘌呤), 下划线的 A 是不变的 (Liao 等, *Virology* 323:131; 2004)。在该序列中, 任何偏差都会引起分支点和剪接受体位点使用效率改变。

在本发明的一个实施例中, 分支点序列加剪接受体序列是: TACTTAT GACTCGTACTATTGTTATTCATCC AG↓G (SEQ ID NO:39)。下划线序列是分支点序列, 箭头表示依照剪接规则的剪接位点位置。因为该规则决定了共有序列并不是不变的, 故其它遵守规则的类似序列也可使用。在一个实施例中, 所使用的分支点加剪接受体位点来自另一种 Ad 血清型。换言之, 分支点加剪接受体位点是异源的, 即并非来自 Ad 载体所基于的天然腺病毒, 但来自不同的 Ad 血清型。

可添加能改变剪接效率和/或期望的转基因表达的其它任选序列。例如, 已鉴定了能增强或抑制剪接的存在于外显子中的顺式作用元件。这些序列称为外显子剪接增强子 (ESE) 或外显子剪接阻抑制子 (ESS) (见 Zheng, *J Biomed Sci* 11:278; 2004 评论)。尽管这些元件没有共有序列, 但已在其它病毒和非病毒生物体中鉴定了许多 ESSs 和 ESEs 的例子, 这些元件都可以使用。本领域技术人员能仔细审视文献和数据, 并选择适当的序列。

总之, 本发明提供了用于将转基因编码区插入在病毒载体基因组特定区域的方法。该方法利用了已知病毒转录元件和表达 Ad 基因的机制, 因为无需额外的启动子且调节信号包含较小尺寸的 DNA 片段, 减小了插入在 Ad 基因组中用于转基因表达的 DNA 序列的尺寸, 提供了转基因时序调节 (如感染早期对晚期; 感染早期对中间期) 的灵活性, 还提供了调节转基因表达量的技术。例如, 通过将转基因插入在通常以高水平表达的转录物中和/或可操作地连接高效剪接受体位点和转基因编码区, 可表达更高数量的转基因。表达水平还受调节信号与其共有序列接近程度的影响; 可进行改变以适应所需表达。

有时, 转基因表达可抑制复制型病毒的生活周期。在这种情况下, 转基因可以仅在感染晚期表达或在主要在感染晚期 (在病毒 DNA 复制

后)表达的方式插入。例如,根据本发明转基因可插入在 L3 中。对于某些转基因,期望在病毒生活周期早期表达转基因。例如,转基因可插入在任一早期区(如 E3)或 L1 区上游。

转录调节元件 (TREs)

本领域已知转录调节元件 (TREs)、以及其鉴定、分离、表征、遗传操作和用于调节可操作地连接的编码序列的方法。TRE 可来源于单基因的转录调节序列,可组合来自不同基因的序列以产生功能性 TRE,或者可通过合成产生 TRE (如 CTP4 启动子)。

取决于靶组织或肿瘤中存在的细胞类型, TRE 可以是组织特异的、肿瘤特异的、发育阶段特异的、细胞状态特异的等。这样的 TREs 在此总的称为组织特异性或靶细胞特异性 TREs。详述如下,靶细胞特异性 TRE 可包含许多构型,包括但不限于靶细胞特异性启动子和靶细胞特异性增强子;异源启动子和靶细胞特异性增强子;靶细胞特异性启动子和异源增强子;异源启动子和异源增强子;以及前述的多聚体。靶细胞特异性 TRE 的启动子和增强子元件可位于目标编码序列的任何方向和/或任何距离,只要获得期望的靶细胞特异性转录活性即可。

可用本领域已知的多种方法(详述如下)测定转录激活,但一般通过检测和/或定量受靶细胞特异性 TRE 控制的(即可操作地连接)编码序列的 mRNA 或蛋白产物来进行测定。

如在此所进一步讨论的,靶细胞特异性 TRE 可具有不同的长度和不同的序列组成。靶细胞特异性 TRE 优先在有限种群(或类型)的细胞中起作用,如前列腺细胞、肝细胞、黑色素瘤细胞等。因此,在某些实施例中,所使用的 TRE 在任一下列组织类型中优先起作用:前列腺;肝;乳腺;泌尿道上皮(膀胱);结肠;肺;卵巢;胰腺;胃;子宫等。

易于为本领域技术人员所理解的是, TRE 是一种多核苷酸序列,且多种序列变换后亦能起作用。本领域已知核苷酸取代、添加和缺失的方法,且易于获得使本领域技术人员能测定序列变体是否显示必需的细胞特异性转录调节功能的功能测定法(例如 CAT 或荧光素酶报道基因测定)。因此,包含核酸取代、添加和/或缺失的保留功能的 TREs 变体可用

于在此披露的载体中。因此，TREs 变体可在靶细胞中保留功能但无须显示最大功能。事实上，对于实现期望结果，并不总是需要 TRE 的最大转录激活活性，对于某些应用，TRE 片段提供的诱导水平就足够。例如，如果用于治疗或缓解疾病状态，如果靶细胞不是特别有毒性的和/或疾病程度相对有限，小于最大的应答性就足够。

某些碱基修饰可产生增加的表达水平和/或细胞特异性。例如，本领域已知 TRE 中核酸序列缺失或添加可使转录调节蛋白结合位点较之正常构型彼此距离更近或更远，或使之旋转以处于 DNA 螺旋的对侧，由此改变 TRE-结合的转录因子之间的空间关系，导致转录减少或增加。因此，尽管希望不受理论束缚，但本发明预计到某些 TRE 修饰会产生改变的处于 TRE 指导下的表达水平的可能性，包括增强的细胞特异性。在更具攻击性的肿瘤生长情况下和/或当要保证更快速和/或更具攻击性地杀死细胞时，特别期望获得增强的表达水平（例如，在无免疫应答的患者中）。

供本发明载体使用的 TRE 可包含或不包含沉默基因。沉默基因（即本领域已知的负调节元件）的存在可帮助关闭非靶细胞中的转录（以及由此的复制）。因此，沉默基因的存在能通过更有效地防止非靶细胞中的复制从而赋予增强的细胞特异性载体复制。或者，缺少沉默基因可刺激靶细胞中的复制，因此赋予增强的靶细胞特异性。

可以多种本领域已知的方法测量 TRE 指导的转录活性（包括抑制和增强），但一般是通过检测和/或定量 mRNA 和/或处于 TRE 控制下（即可操作地连接）的序列编码的蛋白产物来测量。

如本文所讨论的，TRE 可具有不同的长度以及不同的序列组成。在某种程度上，异源 TRE 的尺寸可通过病毒载体容量得以测定，还取决于载体的预期类型。一般来说，TREs 优选具有最低限度的尺寸，其为插入所期望的其它序列提供潜在的空间，例如转基因和/或额外的调节序列。在一个优选的实施例中，这样的额外的调节序列是 IRES、自加工切割序列，例如 2A 或 2A-样序列、剪接序列或分支点。

举例来说，腺病毒载体可用总计达基因组尺寸的约 105%或大约 1.8 kb 的额外序列包装，无需缺失病毒序列。如果从腺病毒基因组上除去非

必需序列，额外的 4.6 kb 插入物就可容许（即总的插入容量为约 6.4 kb）。

就复制型腺病毒载体来说，为了使非特异性复制减到最小，优选从载体上除去内源（腺病毒）TREs（即天然的 E1A 和/或 E1B 启动子）。除利于靶细胞特异性复制之外，除去内源 TREs 还提供了更大的载体插入容量，如果腺病毒载体要包装在病毒颗粒中，则其是特别有意义的。更重要的是，内源 TREs 的缺失阻止了重组事件的可能性，由于异源 TRE 缺失，内源 TRE 承担了其相应的腺病毒编码序列的转录控制（因此容许非特异性复制）。在一个实施例中，构建腺病毒载体使得一个或多个腺病毒基因的内源转录控制序列缺失并为一个或多个异源 TREs 所取代。然而，只要足够的细胞特异性复制优先得以保存，内源 TREs 就可保留在腺病毒载体中。通过在内源 TRE 和编码复制必需区段的基因之间插入异源 TREs 构建了这些实施例。通过比较容许异源 TREs 可起作用的细胞与不起作用的细胞中腺病毒载体复制的传导测定，从而测定了必需的细胞特异性复制优先。

一般来说，与不存在 TRE 下的复制的基础水平相比，TRE 会增加靶细胞中载体复制至少约 2 倍、优选至少约 5 倍、优选至少约 10 倍、更优选至少约 20 倍、更优选至少约 50 倍、更优选至少约 100 倍、更优选至少约 200 倍、还优选至少约 400 至约 500 倍，还优选至少约 1000 倍。可以经验为主地测定可接受的差异（通过使用例如 RNA 印迹测定法、RNase 保护测定法或其它本领域已知的测定法测量 mRNA 水平），并取决于该载体预期的用途和/或所期望的结果。

使用在靶细胞中优先起作用的 TREs 能产生定向于特异性靶细胞的腺病毒载体。在本发明的一个实施例中，靶细胞特异性或细胞状态特异性的异源 TRE 是肿瘤细胞特异性的。载体可包含单个肿瘤细胞特异性的 TRE 或多个肿瘤细胞特异性的并在相同细胞中起作用的异源 TREs。在另一个实施例中，载体包含一个或多个肿瘤细胞特异性的异源 TREs，并额外包含一个或多个组织特异性的异源 TREs，籍此在相同细胞中所有 TREs 都起作用。

在一个优选的溶瘤腺病毒平台实施例中，含有 IRES 或诸如 2A 或 2A-样序列的自加工切割序列的双顺反子或多顺反子盒包含腺病毒早期病毒基因（E1A、E1B、E2、E3 和/或 E4）或在病毒生活周期晚期表达的基因（尾丝、五邻体和六邻体）。

在某些情况下，由于例如癌靶细胞相对难治愈性或特别的攻击性，因此期望提高细胞毒素活性程度和/或等级。有利于细胞毒性的病毒基因的例子包括但不限于腺病毒死亡蛋白（ADP）基因。在本文披露的另一个实施例中，腺病毒包含其内源性启动子缺失或其中有缺失的腺病毒 E1B 基因。在本文披露的其它实施例中，E1B 的 19-kDa 区缺失。

为提供对靶细胞而言增强的细胞毒性，一种或多种具有细胞毒性效应的转基因可存在于载体中。此外，或可选择地，有利于细胞毒性和/或细胞死亡的腺病毒基因，例如腺病毒死亡蛋白（ADP）基因，可包括在载体中，任选地处于异源 TRE 的选择性转录控制下，并任选地处于 IRES 或自加工切割序列例如 2A 或 2A-样序列的转录控制下。可通过将靶细胞特异性细胞毒素活性与异源基因或转基因的细胞特异性表达相结合而得以实现。

任何一种异源治疗基因或转基因都可包括在本发明的复制型病毒载体中，进一步描述于下。

一般来说，前述双顺反子或多顺反子盒被置于处于转录应答元件（通常是优先在癌或肿瘤细胞中表达的与细胞类型或细胞状态有关的转录调节元件）的控制下。因此，取决于靶细胞的类型，包含在给定构建体中的治疗基因会有所变化。

本领域已知 TREs 活性可诱导。在不存在诱导物时，可诱导的 TREs 通常显示低活性，而在诱导物存在下则是向上调节的。诱导物包括例如核酸、多肽、小分子、有机化合物和/或诸如温度、压力或的低氧的环境条件。当期望仅在确定时间或确定位置表达时，或当期望使用诱导剂滴定决定表达水平时，优选可诱导的 TREs。例如，通过雄激素可诱导来自 *PSE-TRE*、*PB-TRE* 和 *hKLK2-TRE* 的转录活性，如本文及在此特别并入作为参考的 PCT/US98/04080 所述。因此，在本发明的一个实施例中，腺病毒载体包含可诱导的异源 TRE。

用于本发明的 TRE 可以多种构型存在。TRE 可包含多聚体。例如，TRE 可包含至少两个、至少三个、至少四个或至少五个靶细胞特异性 TREs 的串联序列。这些多聚体还可含有异源启动子和/或增强子序列。

可选择地，TRE 可包含一个或多个启动子区以及一个或多个增强子区。TRE 多聚体还可包含来自不同基因的启动子和/或增强子序列。TRE 的启动子和增强子元件相对于彼此可以任何方向存在，且可与感兴趣的编码序列以任何方向和/或任何距离存在，只要获得期望的细胞特异性转录活性即可。

本文使用的来源于特定基因的 TRE 与其所来源的基因有关，且是一种调节可操作地连接的多核苷酸序列在表达该基因的宿主细胞中转录的多核苷酸序列。例如，本文使用的“人腺体激肽释放酶转录调节元件”或“*hKLK2*-TRE”是一种多核苷酸序列，优选是一种 DNA 序列，其增加可操作地连接的多核苷酸序列在容许 *hKLK2*-TRE 起作用的宿主细胞中的转录，例如表达雄激素受体的细胞（优选哺乳动物细胞，更优选人细胞），例如前列腺细胞。因此，*hKLK2*-TKE 响应于雄激素受体的结合，并包含至少一部分的 *hKLK2* 启动子和/或 *hKLK2* 增强子（即 ARE 或雄激素受体结合位点）。人腺激素增强子和包含该增强子的腺病毒载体描述于 WO99/06576，在此特别并入作为参考。

本文使用的“probasin (*PB*) 转录调节元件”或“*PB*-TRE”是一种多核苷酸序列，优选是一种 DNA 序列，其选择性增加可操作地连接的多核苷酸序列在容许 *PB*-TRE 起作用的宿主细胞中的转录，例如表达雄激素受体的细胞（优选哺乳动物细胞，更优选人细胞，还优选前列腺细胞）。因此，*PB*-TRE 响应于雄激素受体的结合，并包含至少一部分的 *PB* 启动子和/或 *PB* 增强子（即 ARE 或雄激素受体结合位点）。特异于表达雄激素细胞的腺病毒载体描述于 WO98/39466，在此特别并入作为参考。

本文使用的“前列腺特异性抗原 (*PSA*) 转录调节元件”或“*PSA*-TRE”或“*PSE*-TRE”是一种多核苷酸序列，优选是一种 DNA 序列，其选择性增加可操作地连接的多核苷酸序列在容许 *PSA*-TRE 起作用的宿主细胞中的转录，例如表达雄激素受体的细胞（优选哺乳动物细胞，更优选人细胞，还优选前列腺细胞）。因此，*PSA*-TRE 响应于雄激素受体的结合，并包

含至少一部分的 *PSA* 启动子和/或 *PSA* 增强子（即 ARE 或雄激素受体结合位点）。在前列腺中活化并用于腺病毒载体的组织特异性增强子描述于 WO 95/19434 和 WO 97/01358，其中每篇在此特别并入作为参考。

本文使用的“癌胚抗原（*CEA*）转录调节元件”或“*CEA-TRE*”是一种多核苷酸序列，优选是一种 DNA 序列，其选择性增加可操作地连接的多核苷酸序列在容许 *CEA-TRE* 起作用的宿主细胞中的转录，例如表达 *CEA* 的细胞（优选哺乳动物细胞，更优选人细胞）。*CEA-TRE* 响应于与 *CEA* 生成细胞有关的转录因子和/或辅因子，并包含至少一部分的 *CEA* 启动子和/或增强子。特异于表达癌胚抗原的细胞的腺病毒载体描述于 WO 98/39467，在此特别并入作为参考。

本文使用的“ α -胎蛋白（*AFP*）转录调节元件”或“*AFP-TRE*”是一种多核苷酸序列，优选是一种 DNA 序列，其选择性增加（可操作地连接的多核苷酸序列）在容许 *AFP-TRE* 起作用的宿主细胞中的转录，例如表达 *AFP* 的细胞（优选哺乳动物细胞，更优选人细胞）。*AFP-TRE* 响应于与 *AFP* 生成细胞有关的转录因子和/或辅因子，并包含至少一部分的 *AFP* 启动子和/或增强子。特异于表达 α 胎蛋白的细胞的腺病毒载体描述于 WO 98/39465，在此特别并入作为参考。

本文使用的“黏蛋白基因（*MUC*）转录调节元件”或“*MUC1-TRE*”是一种多核苷酸序列，优选是一种 DNA 序列，其选择性增加（可操作连接的多核苷酸序列）在容许 *MUC1-TRE* 起作用的宿主细胞中的转录，例如表达 *MUC1* 的细胞（优选哺乳动物细胞，更优选人细胞）。*MUC1-TRE* 响应于与 *MUC1* 生成细胞有关的转录因子和/或辅因子，并包含至少一部分的 *MUC1* 启动子和/或增强子。

本文使用的“膀胱上皮细胞特异性转录应答元件”或“膀胱上皮细胞特异性 TRE”是一种多核苷酸序列，优选是一种 DNA 序列，其增加可操作地连接的多核苷酸序列在容许膀胱上皮细胞特异性 TRE 起作用的宿主细胞即靶细胞中的转录。已知多种膀胱上皮细胞特异性 TREs，响应于与膀胱上皮细胞有关的细胞蛋白（转录因子和/或辅因子），并包含至少一部分的膀胱上皮细胞特异性启动子和/或膀胱上皮细胞特异性增强子。示范性的膀胱上皮细胞特异性转录调节序列包括人或啮齿动物尿斑素

(urolakin (UP)) , 如 UPI、UPII、UPIII 等。人膀胱上皮细胞特异性 urolakin 转录调节序列及包含其的腺病毒载体描述于 WO 01/72994, 在此特别并入作为参考。

本文使用的“黑素细胞特异性转录应答元件”或“黑素细胞特异性 TRE”是一种多核苷酸序列, 优选是一种 DNA 序列, 其增加可操作地连接的多核苷酸序列在容许黑素细胞特异性 TRE 起作用的宿主细胞即靶细胞中的转录。已知多种黑素细胞特异性 TREs, 响应于与黑素细胞有关的细胞蛋白(转录因子和/或辅因子), 并包含至少一部分的黑素细胞特异性启动子和/或黑素细胞特异性增强子。在此描述了用于测定黑素细胞特异性 TRE 活性并因此用于测定所给定的细胞是否容许黑素细胞特异性 TRE 起作用的方法。用于实施本发明的黑素细胞特异性 TRE 的例子包括但不限于来源于酪氨酸酶基因 5'侧翼区、酪氨酸酶相关蛋白-1 基因的 TRE, 来源于酪氨酸酶相关蛋白-2 基因 5'侧翼区的 TRE, 来源于 MART-1 基因 5'侧翼区的 TRE 或来源于在黑素瘤中异常表达的基因 5'侧翼区的 TRE。

在另一方面, 本发明提供了包含来源于可操作地连接腺病毒复制必需基因的 PRL-3 基因或转基因的转移性结肠癌特异性 TRE 的腺病毒载体。本文使用的“来源于 PRL-3 基因的转移性结肠癌特异性 TRE”或“PRL-3 TRE”是一种多核苷酸序列, 优选是一种 DNA 序列, 其选择性增加可操作地连接的多核苷酸序列在容许 PRL-3 TRE 起作用的宿主细胞中的转录, 例如细胞(优选哺乳动物细胞、更优选人细胞、还优选转移性结肠癌细胞)。转移性结肠癌特异性 TRE 可包含一个或多个调节序列, 如增强子、启动子、转录因子结合位点等, 其可来源于相同的或不同的基因。在一个优选的方面, PRL-3 TRE 包含 PRL-3 启动子。已发现在转移性结肠癌中 PRL-3 蛋白酪氨酸磷酸酶基因以高水平特异性表达(Saha 等(2001) Science 294:1343)。最初鉴定属于转移性结肠癌文库上调节基因组, 通过基因表达系列分析(SAGE)鉴定后, 证实 PRL-3 基因仅在转移灶中而不在原发癌或癌变前腺瘤中升高。复制型腺病毒载体包含描述于 WO 20004/009790 的 PRL-3 转录调节序列。相关序列的例子是存在于 PRL-3 基因翻译起始密码子上游 0.6 kb 和 1 kb 序列(在 WO 20004/009790 中称为 SEQ ID NO:1 和 SEQ ID NO:2)。

在另一方面，本发明提供了包含来源于可操作地连接腺病毒复制必需基因或转基因的 CRG-L2 基因的肝癌特异性 TREs 腺病毒载体。本文使用的“来源于 CRG-L2 基因的肝癌特异性 TREs”或“CRG-L2 TRE”是一种多核苷酸序列，优选是一种 DNA 序列，其选择性增加可操作地连接的多核苷酸序列在容许 CRG-L2 起作用的宿主细胞中的转录，例如细胞（优选哺乳动物细胞、更优选人细胞、还优选肝癌细胞）。肝细胞癌特异性 TRE 可包含一个或多个调节序列，如增强子、启动子、转录因子结合位点等，其可来源于相同的或不同的基因。在一个优选的方面，CRG-L2 TRE 可来源于 CRG-L2 基因翻译起始密码子上游 0.8 kb 序列，或来源于包含于该 0.8 kb 序列中的 0.7 kb 序列（残基 119-803）；或来源于该 0.8 kb 序列的 EcoRI 至 NcoI 片段，其描述于序列号 60/511,812 美国临时申请，在此特别并入作为参考。

在另一方面，本发明提供了包含可操作地连接腺病毒复制必需基因或转基因的 EBV-特异性转录调节元件（TRE）的腺病毒载体。在一方面，EBV 特异性 TRE 来源于 LMP1、LMP2A 或 LMP2B 基因翻译起始密码子上游的序列，详述于美国临时申请序列号 60/423,203，在此特别并入作为参考。该 EBV-特异性 TRE 可包含一个或多个调节序列，如增强子、启动子、转录因子结合位点等，其可来源于相同的或不同的基因。

在又一方面，本发明提供了包含可操作地连接腺病毒复制必需基因或转基因的缺氧应答元件（“HRE”）腺病毒载体。HRE 是一种包含转录复合物 HIF-1 结合位点或缺氧诱导因子-1 的转录调节元件，其与包括血管内皮生长因子的几个基因的调节区相互作用，或与编码包括烯醇酶-1 的糖分解酶类的几个基因的调节区相互作用。因此，在一个实施例中，腺病毒载体包含处于诸如 HRE 的细胞状态特异性 TRE 转录控制下的腺病毒基因，优选复制必需的腺病毒基因，详述于 WO 00/15820，在此特别并入作为参考。

另一方面，本发明提供了包含可操作地连接腺病毒复制必需基因或转基因的“端粒酶启动子”或“TERT 启动子”的腺病毒载体。本文使用的术语“端粒酶启动子”或“TERT 启动子”指天然 TERT 启动子及其功能性片段、变体和衍生物。TERT 启动子不必是全长或野生型启动子。本领域技

术人员熟知如何由 TERT 启动子获得片段并测试它们所期望的选择性。本发明的 TERT 启动子片段具有肿瘤细胞选择性启动子活性，即驱动肿瘤选择性表达可操作地连接的编码序列。在一个实施例中，本发明的 TERT 启动子是哺乳动物 TERT 启动子。在另一个实施例中，哺乳动物 TERT 启动子是人 TERT (hTERT) 启动子。对于供本发明组合物和方法使用的示范性的 TERT 启动子，参见如 WO 98/14593 和 WO 00/46355。

在又一方面，本发明提供了包含可操作地连接腺病毒复制必需基因或转基因的“E2F 启动子”的腺病毒载体。本文使用的术语“E2F 启动子”指天然 E2F 启动子及其功能性片段、变体和衍生物。E2F 启动子不必是全长或野生型启动子。本领域技术人员熟知如何由 E2F 启动子获得片段并测试它们所期望的选择性。本发明的 E2F 启动子片段具有肿瘤细胞选择性启动子活性，即驱动肿瘤选择性表达可操作地连接的编码序列。本文使用的术语“肿瘤选择性启动子活性”指本发明的启动子片段的启动子活性在肿瘤细胞中高于在非肿瘤细胞类型中。E2F-应答启动子具有至少一个 E2F 结合位点。在一个实施例中，E2F-应答启动子是哺乳动物 E2F 启动子。在另一个实施例中，其是人 E2F 启动子。例如，E2F 启动子可以是人 E2F-1 启动子。此外，人 E2F-1 启动子可以是例如具有 SEQ ID NO:43 所述序列的 E2F-1 启动子。本领域已知多个 E2F 启动子的例子（如 Parr 等 *Nature Medicine* 1997:3 (10) 1145-1149, WO 02/067861, US20010053352 和 WO 98/13508）。E2F 应答启动子通常具有共同的特征，例如邻近其 E2F 位点的 Sp I 和/或 ATT7 位点，该 E2F 位点常位于转录起始位点附近，还缺少可识别的 TATA 盒。E2F-应答启动子包括 E2F 启动子，例如 E2F-1 启动子、二氢叶酸还原酶 (DHFR) 启动子、DNA 聚合酶 A (DPA) 启动子、c-myc 启动子和 B-myb 启动子。E2F-1 启动子含有四个在血清饥饿 (serum-starved) 细胞中作为转录阻遏元件的 E2F 位点。在一个实施例中，E2F-应答启动子具有至少两个 E2F 位点。在另一个实施例中，E2F 启动子可操作地连接腺病毒 E1a 区。在又一个实施例中，E2F 启动子可操作地连接腺病毒 E1b 区。在再一个实施例中，E2F 启动子可操作地连接腺病毒 E4 区。

在本发明的一个实施例中，本发明的重组病毒载体在 Rb-通道缺陷细胞中选择性复制并溶解该细胞。在一个实施例中，本发明的 E2F 启动子是哺乳动物 E2F 启动子。在另一个实施例中，哺乳动物 E2F 启动子是人 E2F 启动子，例如包含或基本上由 SEQ ID NO:43 组成的人 E2F 启动子。本发明的实施例包括包含 E2F 启动子的腺病毒载体，其中 E2F 启动子包含的核苷酸序列选自：（a）SEQ ID NO:43 所示的核苷酸序列；（b）SEQ ID NO:43 所示的核苷酸序列的片段，其中该片段具有肿瘤选择性启动子活性；（c）在其全长上与 SEQ ID NO:43 所示的核苷酸序列具有至少 90、91、92、93、94、95、96、97、98、99%或更多%同一性的核苷酸序列，其中该核苷酸序列具有肿瘤选择性启动子活性；以及（d）在严谨条件下与 SEQ ED NO:43 所示的序列杂交的具有全长互补的核苷酸序列，其中该核苷酸序列具有肿瘤选择性启动子活性。在本发明的另一个实施例中，E2F 启动子包含 SEQ ID NO:43 第 7-270 位核苷酸。在本发明的另一个实施例中，E2F 启动子包含 SEQ ID NO:43 第 7-270 位核苷酸，其中 SEQ ID NO: 43 第 75 位核苷酸是 T 而不是 C。

在另一个实施例中，当比较并比对最大对应时，如使用以下序列比较算法或通过目视检查所测定的，根据本发明的 E2F 启动子与 SEQ ID NO:43 所示的核苷酸序列具有至少 80、85、87、89、90、91、92、93、94、95、96、97、98、99%或更多的序列同一性。在一个实施例中，所给定的%序列同一性存在于至少约 50 个核苷酸长的序列的区中。在另一个实施例中，所给定的%序列同一性存在于至少约 100 个核苷酸长的序列的区中。在另一个实施例中，所给定的%序列同一性存在于至少约 200 个核苷酸长的序列的区中。在另一个实施例中，所给定的%序列同一性存在于全长序列中。

E2F-应答启动子不必是全长的或野生型启动子，但应具有至少 3 倍、至少 5 倍、至少 10 倍、至少 20 倍、至少 30 倍、至少 50 倍、至少 100 倍或甚至至少 300 倍的肿瘤选择性。可通过多种测定法测定肿瘤选择性，所述测定法使用了诸如 WO 02/067861 实施例 4 中使用的技术的已知技术，例如 RT-PCR 或在所选择的细胞类型中比较复制。

不受理论束缚，E2F-应答启动子（以下有时简称 E2F 启动子）的选择性据报道是基于 Rb-通道缺陷肿瘤细胞中 E2F 启动子/反式作用子的脱阻遏。在休眠细胞中，E2F 结合三元复合物中的肿瘤抑制蛋白 pRB

蛋白尿激酶纤维蛋白溶酶原激活剂（uPA）及其细胞表面受体尿激酶纤维蛋白溶酶原激活剂受体（uPAR）表达于许多最常见的赘生物中，似乎是癌转移中重要的蛋白。这两种蛋白都涉及乳腺癌、结肠癌、前列腺癌、肝癌、肾癌、肺癌和卵巢癌。调节 uPA 和 uPAR 转录的序列元件已经被广泛地研究（Ricchio 等（1985）*Nucleic Acids Res.* 13:2759-2771; Cannio 等（1991）*Nucleic Acids Res.* 19:2303-2308; 也参见 WO 98/39464）。

可操作地连接必需 Ad 编码区的异源 TRE(s)

对于早期基因的操作，Ad5 E1A 的转录起始位点在病毒基因组的第 498 位，E1A 编码区段的 ATG 起始位点在病毒基因组的第 560 位。该区可用于插入异源 TRE。图 1 描述了 Ad5 天然基因组结构，图 2 描述了天然 Ad5 转录单位。

通过使用聚合酶链式反应（PCR）可导入受限酶切位点，在此所使用的引物限于 Ad5 基因组，或可包含携带 Ad5 基因组 DNA 的质粒的一部分。例如，在使用 pBR322 时，引物可使用 pBR322 骨架中的 EcoRI 位点和 Ad5 nt 1339 处的 XbaI 位点。PCR 分两步进行，在该区中心的重叠引物处导入核苷酸序列改变，产生独特的限制酶切位点，可供在该位点插入异源 TRE 之用。

类似的策略亦可用于将异源 TRE 元件插入在可操作连接的 E1B 中。Ad5 的 E1B 启动子由 Sp1 的单个高亲和力识别位点以及 TATA 盒组成。该区从 Ad5 nt 1636 延伸到 1701。通过在该区中插入细胞特异性异源 TRE，可供 E1B 基因细胞特异性转录之用。通过使用调节 E1A 的细胞特异性应答元件修饰的左侧区，作为模板用于导入异源 TRE 以调节 E1B，所得到的腺病毒载体将依赖于细胞特异性转录因子用于表达 E1A 和 E1B。在一些实施例中，E1B 的部分或全部 19-kDa 区缺失。

同样地，异源 TRE 可插入在 E2 基因上游使其细胞特异性表达。在 Ad5 中定位于约 27050-27150 的 E2 早期启动子由主要和次要转录起始位点，后者占 E2 转录物的约 5%，2 个非规范的 TATA 盒，两个 E2F 转录因子结合位点和 ATF 转录因子结合位点组成（E2 启动子结构的详细论述见 Swaminathan 等, *Curr. Topics in Micro, and Immunol.* (1995) 199 (part 3) : 177- 194)。

E2 晚期启动子与相反链编码的基因的编码序列重叠，因此无法进行遗传操作。然而，E2 早期启动子与相反链上编码 33 kD 蛋白的序列仅重叠几个碱基对。值得注意的是 SpeI 限制酶切位点（Ad5 第 27082 位）是上述 33 kD 蛋白终止密码子的一部分，能方便地将主要 E2 早期转录起始位点和 TATA-结合蛋白位点与从上游转录因子结合位点 E2F 和 ATF 分离。因此，在 SpeI 位点内插入具有 SpeI 末端的异源 TRE 会破坏 Ad5 的内源性 E2 早期启动子，且应容许 E2 转录物的细胞特异性表达。对于 E4，可使用腺病毒基因组的右侧部分。对于 Ad5，E4 转录起始位点主要在约 nt 35605，TATA 盒在约 nt 35631 以及 ORF I 的第一 AUG/CUG 在约 nt 35532。Virtanen 等（1984）*J. Virol.* 51: 822-831。对于其它基因使用任何一种上述策略，异源 TRE 可被导入转录物起始位点上游。为了构建具有插入在 E4 区的异源 TRE 的全长腺病毒，可在 W162 细胞（Weinberg 等（1983）*Proc. Natl. Acad. Sci.* 80:5383-5386）中进行共转染和同源重组，其提供了 E4 蛋白以反向补偿这些蛋白合成中的缺陷。

“E3 区”（与“E3”可交换使用）是本领域众所周知的术语，指编码 E3 基因产物的腺病毒基因组的区。在各种出版物中已描述了 E3 区，包括例如，Wold 等（1995）*Curr. Topics Microbiol. Immunol.* 199:237-274。E3 区的“一部分”指小于完整的 E3 区，例如包括多核苷酸缺失以及编码一个或多个 E3 区多肽产物的多核苷酸。见图 6、7 和 8。

在含有 E3 的腺病毒质粒例如 BHGE3（Microbix Biosystems Inc., Toronto）和不含有 E3 的腺病毒质粒之间进行同源重组可产生含有 E3 区的腺病毒构建体。

可选择地，可将包含 E3 区的腺病毒载体与腺病毒构建体或腺病毒质粒构建体一起导入细胞，例如 293 细胞，在此它们能经历同源重组以产生

含有 E3 区的腺病毒。在该情况下，含有 E3 的腺病毒载体和腺病毒构建体或质粒构建体含有腺病毒的互补区，例如一种含有左侧区，而其它的则含有右侧区，具有足够的序列重叠以便进行同源重组。

可选择地，可使用其它常规方法构建本发明含有 E3 的腺病毒载体，包括标准重组法（如使用受限核酸酶和/或 PCR）、化学合成或这些方法的组合。此外，使用标准的分子生物学技术可进行 E3 区一部分的缺失。

在某些实施例中，E3 区中编码的腺病毒死亡蛋白（ADP）被保留在腺病毒载体中。该 ADP 基因受主要晚期启动子（MLP）控制，编码对于加速宿主细胞溶解起重要作用的蛋白（ADP）。Tollefson 等（1996）*J. Virol.* 70（4）:2296；Tollefson 等（1992）*J. Virol.* 66（6）:3633。因此，含有 ADP 基因的腺病毒载体可使腺病毒载体更有效，使更有效地治疗和/或更低的剂量需要成为可能。

因此，在一个实施例中，本发明提供了其中腺病毒基因处于第一 TRE 转录控制下以及编码 ADP 的多核苷酸序列处于第二 TRE 元件控制下的腺病毒载体，其中优选地腺病毒基因是复制必需的。编码 ADP 的 DNA 序列和 ADP 的氨基酸序列是公众可得到的。简言之，使用本领域已知技术例如 PCR 由 Ad 获得 ADP 编码序列。优选地，还获得 Y 前导区（其是一种用于晚期基因正确表达的重要序列）并连接 ADP 编码序列。然后，ADP 编码序列（具有或不具有 Y 前导区）可导入腺病毒基因组中，例如在 E3 区中（在此 ADP 编码序列由 MLP 启动）。ADP 编码序列还能插入在腺病毒基因组的其它位置，例如 E4 区。可选择地，ADP 编码序列能可操作地连接不同类型的 TRE，包括但并不仅限于其它病毒 TRE。在一个实施例中，本发明的载体具有可操作地连接其天然 TREs 的 ADP。

内部核糖体进入位点

为由一个病毒或非病毒载体表达两个或更多个蛋白，通常使用内部核糖体进入位点（IRES）序列来驱动第二、第三、第四个基因等的表达。本发明的腺病毒载体可包含一个或多个基因间 IRES 元件，其连接着两个或更多个编码序列的翻译。包含连接两个腺病毒编码区的 IRES 的腺病毒载体是稳定的，且与不含有 IRES 的载体相比可提供更好的特异性。

包含基因间 IRES 的腺病毒载体而不是第二 TRE 则可以提供额外的载体空间用于包含额外的基因如治疗基因。包含 IRES 的腺病毒载体的例子描述于 US 6,692,736, 在此特别并入作为参考。在本发明一方面中, 病毒载体在多顺反子转录物中包含至少一个 IRES, 其中多顺反子转录物的产生受异源、细胞型、组织型或细胞状态特异性 TRE 的调节。对于包含处于 IRES 控制下的第二腺病毒编码区的腺病毒载体, 优选处于 IRES 翻译控制下的编码区的内源启动子缺失, 以便内源启动子不干扰第二编码区转录。如果 IRES 含有起始密码子, 优选第二编码区与 IRES 在读框内。如果 IRES 中存在诸如 ATG 的起始密码子, 则优选除去第二编码序列的起始密码子, 且 IRES 和第二编码区在读框内。或者, 如果 IRES 不含有起始密码子或如果起始密码子从 IRES 上除去, 则使用第二编码区的起始密码子。在一个实施例中, 腺病毒载体包含腺病毒必需基因(处于异源 TRE 转录控制下的 E1A 和 E1B 基因)以及导入在 E1A 和 E1B 之间的 IRES。因此, E1A 和 E1B 处于共同的转录控制之下, 由于 IRES 的存在获得了 E1B 编码区的翻译。在一个实施例中, E1A 的内源性启动子缺失。在另一个实施例中, E1A 的内源性增强子缺失, 还在另外的实施例中, E1A 的内源性启动子和增强子都缺失。在另一个实施例中, E1B 的内源性启动子缺失。还在另外的实施例中, E1B 的 19-kDa 区部分或全部缺失。

利用本领域已知以及本文之前所描述的方法和技术实现将 IRES 插入载体, 包括但不限于限制性内切酶消化、连接和 PCR。可通过化学合成或通过制备例如小 RNA 病毒 IRES 的 cDNA 副本, 获得 IRES 的 DNA 副本。参见例如, Duke 等 (1995) *J.Virol.* 66 (3) :1602-9) 对 EMCV IRES 的描述和 Huez 等 (1998), *Mol. Cell. Biol.* 18 (11) :6178-90) 对 VEGF IRES 的描述。在使其位于多顺反子 mRNA 中 5'-远侧编码区上游的位点, 将内部翻译起始序列插入载体基因组。例如, 在一个实施例中, 其中腺病毒载体产生的双顺反子 E1A-E1B mRNA 处于异源 TRE 的控制下, E1B 启动子缺失或失活, IRES 序列置于 E1A 和 E1B 之间。在另一个实施例中, E1B 的 19- kDa 区部分或全部缺失。心病毒和某些口蹄疫病毒的 IRES 序列在其 3'末端含有 AUG 密码子, 其既作为核糖体进入位点又作为翻译起始位点。因此, 将这种类型的 IRES 导入载体以便替换其调节翻译

的蛋白的翻译起始密码子。然而，在肠道/鼻病毒组的 IRES 中，IRES 3'末端的 AUG 仅用于核糖体进入，翻译是邻接的下游 AUG 密码子启动的。因此，如果肠道/鼻病毒 IRES 在载体中用于下游编码区的翻译调节，则下游基因的 AUG（或其它翻译起始密码子）保留在载体构建体中。

在另一个实施例中，IRES 可操作地连接到插入在并不紧接 polyA 信号上游的腺病毒 CDS 下游的转基因。例如，在前导区序列中，在最下游 Ad CDS 之后并不插入 IRES 转基因。例如，IRES 转基因盒可操作地连接如下 Ad CDSs 之一：52/55K、pV、五邻体、pVI 或六邻体。

多个编码序列能与 IRESs 相连接。在一个实施例中，Ad CDS 通过第一 IRES 可操作地连接第一转基因，所述第一转基因通过第二 IRES 可操作地连接第二转基因。在一个实施例中，第一和第二转基因编码相同的或不同的蛋白。就相同的蛋白来说，有利之处在于转基因之一的编码序列可以是“再编码的”。换言之，使用不同的密码子编码相同的氨基酸。可在 DNA 水平上减少两个转基因之间的同源性数值，由此减少或消除两个转基因之间的同源重组。其它的实施例包括通过 IRES 可操作地连接的两个腺病毒 CDSs。可伴随有腺病毒 DNA 序列的缺失。例如，位于相同前导区中且毗连的两个腺病毒 CDSs 可通过 IRES 可操作地连接，介于其间的 Ad 序列可部分或全部缺失，只要缺失不破坏病毒载体产生必需的其它序列或元件即可，尤其要注意互补链。所缺失的部分可以是 1-5 个核苷酸（nts）、6-15 nts、16-25 nts、26-35 nts、36-40 nts 或大于 40 nts。在一个实施例中，第一转基因 CDS 通过第一 IRES 可操作地连接 Ad CDS，第二 IRES 将第二转基因与所述 Ad CDS 连接。其它实施例包括多种 Ad CDSs 以及 Ad CDSs 和与 IRES 和/或自加工肽序列可操作地连接的转基因 CDSs 的组合。

当在载体中使用多个 IRES 序列时，优选两个 IRES 序列在 DNA 水平上具有最小同源性或不具同源性，以减少同源重组的频率。

自加工切割位点或序列

在本发明另一方面，“自加工切割位点”（如 2A-样序列）被用来表达来自一个 mRNA 的两种多肽。“自加工切割位点”或“自加工切割序列”定义

为一种 DNA 或氨基酸序列，其中在翻译后，发生包含自加工切割位点的多肽的快速分子内（顺式）切割，导致分离的成熟蛋白或多肽产物表达。这样的“自加工切割位点”也可称为翻译后或共翻译加工切割位点，在此例子是 2A 位点、序列或结构域。本文使用的“自加工肽”在此定义为编码自加工切割位点或序列的 DNA 序列的肽表达产物，其翻译后，介导包含自加工切割位点的蛋白或多肽快速分子内（顺式）切割以产生不连续的成熟蛋白或多肽产物。已报道 2A 位点、序列或结构域通过修饰核糖体活性以促进酯键水解，由此从多肽中以容许进行分离的下游翻译产物合成的方式释放翻译复合物，从而证实其翻译效应（Donnelly 等 *J Gen Virol.* 2001 年 5 月; 82 (Pt 5) : 1013-25)。可选择地，还报道了 2A 位点、序列或结构域通过顺式切割其自身 C-末端以产生初级切割产物，从而证实其“自体溶解”或“切割”（Furler; Palmenberg, *Ann. Rev. Microbiol.* 44:603-623 (1990)）。

尽管该机制不是本发明的一部分，但是 2A-样序列的活性可涉及密码子之间的核糖体跳跃（skipping），这可以阻止肽键的形成（de Felipe 等, *Human Gene Therapy* 11:1921-1931 (2000)；Donnelly 等, *J. Gen. Virol.* 82:1013-1025 (2001)；Donnelly 等 *J Gen Virol.* 2001 May;82 (Pt 5) : 1027- 41)；Szymczak 等 *Nature Biotechnology* 22:589-594 和 760 (2004)，尽管已考虑了该结构域作用更类似于一种自溶酶（Ryan 等, *Virol.* 173:35-45 (1989)）。对克隆入表达载体并转染靶细胞的口蹄疫病毒（FMDV）2A 编码区的研究显示 FMDV 2A 切割小麦胚裂解物和转基因烟草植物（Halpin 等, *US 5,846,767*; 1998 和 Halpin 等, *Plant J* 17:453-459, 1999)；Hs 683 人神经胶质瘤细胞系（de Felipe 等, *Gene Therapy* 6:198-208, 1999)；以下简称“de Felipe II”）；兔网织红细胞裂解物和人 HTK-143 细胞（Ryan 等, *EMBO J.* 13:928-933 (1994)）；以及昆虫细胞（Roosien 等, *J. Gen. Virol.* 71:1703-1711, 1990）中的人工报道分子多蛋白。对于 IL-12，已显示 FMDV 2A-介导的异源多蛋白切割（p40/p35 异二聚体：Chaplin 等, *J. Interferon Cytokine Res.* 19:235-241, 1999）。该参考文献证实了在转染的 COS-7 细胞中，FMDV 2A 介导 p40-2A-p35 多蛋白切割为具有与 IL-12 有关活性的生物学功能亚基 p40 和 p35。

FMDV 2A 序列已单独或与不同的 IRES 序列组合掺入逆转录病毒载体中，以构建双顺反子、三顺反子和四顺反子载体。Furler 等 (Gene Ther. 2001 年 6 月;8 (11) :864-73) 通过使用编码 α -突触核蛋白和 EGFP 或 Cu/Zn 超氧化物歧化酶 (SOD-1) 和通过 FMDV 2A 序列连接的 EGFP 的重组腺病毒伴随病毒 (AAV) 载体证实了 2A-介导的基因在动物中的表达效率。相对于相应的基于 IRES 的载体，在包括 2A 序列的载体中 EGFP 和 α -突触核蛋白表达水平实质上更高，而 SOD-1 表达水平则相当或稍高。Furler 还证实了在大鼠黑质注射了含有 2A 的 AAV 载体后，2A 序列在体内产生双顺反子基因表达。Szymczak 等 (Nature Biotechnology 22:589-594&760 (2004)) 描述了具有用三个 2A 序列连接四个编码区的逆转录病毒载体。

对于本发明，来源于小 RNA 病毒，包括但不限于肠道病毒、鼻病毒、心病毒、口蹄病毒或口蹄疫病毒 (FMDV) 的病毒序列列举说明了编码自加工切割位点的 DNA 序列。在一个优选的实施例中，自加工切割位点编码序列来源于 FMDV。自加工切割位点包括但不限于 2A 和 2A-样位点、序列或域 (Donnelly 等, J. Gen. Virol. 82:1027-1041 (2001))。

FMDV 2A 是一种多蛋白区，其在 FMDV 基因组中起指导在其自身 C-末端进行单一切割的作用，从而顺式起作用。FMDV 2A 结构域通常报道有约 19 个氨基酸长度 ((LLNFDLLKLAGDVESNPGP (SEQ ID NO:1) ; TLNFDLLKLAGD VESNPGP (SEQ ID NO:2) ; Ryan 等, J. Gen. Virol. 72:2727-2732 (1991)))，然而少至 14 个氨基酸残基的寡肽 ((LLKLAGDVESNPGP (SEQ ID NO:3))) 也能显示以类似于其在天然 FMDV 多蛋白加工中角色的方式介导在 2A C-末端的切割。已研究了 2A 序列变体介导多蛋白有效加工的能力 (Donnelly 等, J. Gen. Virol. 82:1027-1041 (2001))。2A 序列同源物和变体也包括在本发明的范围内，包括但不限于如 SEQ ID NOs: 1-32 所示的序列。

在一个实施例中，包括在根据本发明的载体中的 FMDV 2A 序列编码包含如 SEQ ID NO:1 所示序列的氨基酸残基。可选地，根据本发明的载体可编码如 Donnelly 等, J. Gen. Virol. 82:1027-1041 (2001) 中所讨论的其它 2A-样区的氨基酸残基，其包括但不限于来自小 RNA 病毒、昆虫病

毒、C型轮状病毒、锥虫重复序列或细菌、热海栖热袍菌（*Thermatoga maritima*）的2A-样结构域。

本发明涉及编码诸如2A或2A-样多肽的自加工切割位点的核酸序列变体，以及相对于亲本（天然）核苷酸，一个或多个氨基酸具有不同密码子的核酸编码序列的应用。这样的变体是本发明所特别涉及并包含的。自加工切割肽和多肽的序列变体也包括在本发明的范围内。

依照本发明，还包含的是编码自加工切割多肽和自身与天然序列具有80、85、88、89、90、91、92、93、94、95、96、97、98、99%或更多的序列同一性的多肽的序列变体。

在本发明的一个实施例中，自加工切割序列（如2A或2A-样序列）可操作地连接腺病毒蛋白编码区和转基因。腺病毒蛋白CDS可位于自加工切割位点上游，转基因位于下游。可选择地，转基因CDS可位于自加工切割位点上游，而腺病毒蛋白CDS位于下游。

多个CDSs可连接有自加工切割位点。在一个实施例中，Ad CDS通过自加工切割位点可操作地连接第一转基因，所述第一转基因通过自加工切割位点可操作地连接第二转基因。在一个实施例中，第一和第二转基因编码相同或不同的蛋白。就相同的蛋白来说，有利之处在于转基因之一的编码序列可是“再编码的”。换言之，使用不同的密码子编码相同的氨基酸。可在DNA水平上减少两个转基因之间的同源性数值，由此减少或消除两个转基因之间的同源重组。其它的实施例包括通过自加工切割位点可操作地连接的两个Ad CDSs。可伴随有腺病毒序列的缺失。例如，位于相同前导区中且毗连的两个腺病毒CDSs可通过自加工切割位点可操作地连接，介于其间的Ad序列可部分或全部缺失，只要缺失不破坏病毒载体产生必需的其它序列或元件即可，尤其要注意互补链。所缺失的部分可以是1-5个核苷酸（nts）、6-15 nts、16-25 nts、26-35 nts、36-40 nts或大于40 nts。

在一个实施例中，通过第一自加工切割位点，第一转基因CDS可操作地连接Ad CDS，通过第二自加工切割位点，Ad CDS可操作地连接第二转基因。其它实施例包括多种Ad CDSs以及Ad CDSs和与IRES和/或自加工肽序列可操作地连接的转基因CDSs的组合。

当在载体中使用多于一个的自加工肽序列时，优选自加工肽序列在 DNA 水平上具有最小同源或不具有同源性，以减少同源重组频率。例如自加工肽序列可来源于不同来源，其中多个自加工肽序列的编码序列具有最小同源或不具有同源性。在另一个实施例中，自加工肽序列的编码序列被再编码。换言之，不同的密码子被用于编码自加工肽序列的相同氨基酸。这样做可减少多于一个自加工肽编码序列之间的同源性数值，因此减少了或消除了两个转基因之间的同源重组。

当自加工肽序列的编码序列插入具有上游和下游 CDS 的读框中时，自加工肽序列可操作地连接 CDS。

自加工肽序列的移除

人们所关心的是与使用自加工肽有关的，例如 2A 或 2A-样序列，所述 2A-样序列是含有源自自加工肽的氨基酸的表达多肽的 C 末端，即 2A-衍生的氨基酸残基。对于宿主这些氨基酸残基是“外源的”，当重组蛋白体内表达或体外或离体表达后体内转移时，可引起免疫应答。此外，如果不被切除，自加工肽衍生的氨基酸残基可干扰蛋白功能和/或改变蛋白构象，导致达不到最佳表达水平和/或重组蛋白生物活性减少。换言之，取决于应用，其有利之处在于所得到的蛋白不含有所有 2A-衍生的氨基酸残基。

本发明包括改造的载体，使得在第一蛋白或多肽编码序列（第一或 5' ORF）和自加工切割位点之间提供了额外的蛋白水解切割位点，作为一种手段用于移除衍生自存在于表达的蛋白产物中的氨基酸残基的自加工切割位点。

额外的蛋白水解切割位点的例子是具有共有序列 RXK(R)R（SEQ ID NO:33）的弗林蛋白酶切割位点，其能为内源枯草杆菌蛋白酶样蛋白酶所切割，例如弗林蛋白酶和其它丝氨酸蛋白酶。如 USSN 10/831304 的实施例 6 所示，发明人已阐明通过在第一多肽和自加工 2A 序列之间导入弗林蛋白酶切割位点 RAKR（SEQ ID NO:33），可有效地除去在第一表达蛋白 C 末端的自加工 2A 氨基酸残基。此外，与使用仅含有 2A 序列的质粒相比，使用含有 2A 序列以及与 2A 序列邻接的弗林蛋白酶切割位点的质

粒产生更高水平蛋白表达。这种改善提供了更多的优点，即当 2A 氨基酸残基从蛋白的 C-末端除去时，可使用更长的 2A-或 2A 样序列或其它自加工序列，如 USSN 10/831304 所描述，在此特别并入作为参考。

如在此详述的，2A 肽序列提供了在翻译加工过程中利于免疫球蛋白两条链或其它蛋白产生的“切割”位点。在一个示范性的实施例中，诸如免疫球蛋白重链的第一蛋白的 C-末端含有来源于 2A 序列自身的约 13 个氨基酸残基。残留氨基酸的数量取决于所使用的 2A 序列。如上述及实施例所示，当弗林蛋白酶切割位点序列(如 RAKR)插入在第一蛋白和 2A 序列之间时，2A 残基被从第一蛋白的 C-末端除去。然而，质谱数据显示 RAKR-2A 构建体表达的第一蛋白的 C-末端含有两个来源于弗林蛋白酶切割位点 RAKR 的额外的氨基酸残基 RA。

在一个实施例中，本发明提供了一种用于除去这些残留氨基酸的方法以及用于表达其的组合物。已设计了许多新的构建体，其提供用于从蛋白的 C-末端除去这些额外的氨基酸。弗林蛋白酶切割发生在切割位点的 C-末端，该切割位点具有共有序列 RXR(K)R，其中 X 是任何氨基酸。在一方面，本发明提供了一种用于从蛋白的 C-末端除去新暴露的碱性氨基酸残基 R 或 K 的方法，利用了称为羧肽酶 (CPs) 的一组酶，其包括但不限于羧肽酶 D、E 和 H (CPD、CPE、CPH)。由于 CPs 能在蛋白 C-末端除去碱性氨基酸残基，所有来源于只含有碱性氨基酸 R 或 K 例如 RKKR、RKRR, RRRR 等的弗林蛋白酶切割位点的氨基酸残基能通过 CP 切除。已构建了一系列含有 2A 序列和在 C 末端具有碱性氨基酸残基的弗林蛋白酶切割位点的免疫球蛋白表达构建体以评估切割和残基切除效率。示范性的构建体设计如下：H 链-弗林蛋白酶（如 RKKR、RKRR、RRKR 或 RRRR）-2A-L 链或者 L 链-弗林蛋白酶（如 RKKR、RKRR、RRKR 或 RRRR）-2A-H 链。

本领域技术人员应当清楚，免疫球蛋白重 (H) 链的 C 末端存在碱性氨基酸残基 (K) (使其适合于羧肽酶切割)，而免疫球蛋白轻 (L) 链末端则是非碱性氨基酸 C。在本发明一个优选的实施例中，提供了包含弗林蛋白酶位点和 2A 序列的抗体表达构建体，其中免疫球蛋白 L 链位于免

疫球蛋白 H 链的 5'，以便在翻译后使用羧肽酶切除额外的弗林蛋白酶氨基酸残基。

产生具有完全人特性的治疗性蛋白、多肽、其片段或类似物通常是有利的。这些试剂避免了源自不同物种的蛋白、多肽、其片段或类似物所诱导的不期望的免疫应答。为了引导可能的宿主对来源于自加工肽的氨基酸残基的免疫应答，可将蛋白水解位点编码序列插入在（使用本领域已知的标准方法）第一蛋白编码序列和自加工肽编码序列之间，以便从表达的蛋白或多肽的上除去自加工肽序列。

可被重组 DNA 技术表达任何一种本领域已知的额外的蛋白水解切割位点可被用于实施本专利。示范性的能插入在多肽或蛋白编码序列和自加工切割序列之间的额外的蛋白水解切割位点包括但不限于：

- a) 弗林蛋白酶共有序列或位点：RXK(R)R (SEQ ID NO:33)；
- b) 因子 Xa 切割序列或位点：IE(D)GR (SEQ ID NO:34)；
- c) 信号肽酶 I 切割序列或位点：如 LAGFATVAQA (SEQ ID NO:35)；和
- d) 凝血酶切割序列或位点：LVPRGS (SEQ ID NO:36)；
- e) 腺病毒共有蛋白酶序列或位点(M,L,I)XGG/X (SEQ ID NO:37) 和 (M,L,I)XGX/G (SEQ ID NO:38)，参见 Webster 等 J Gen Virol 70:3215-3223 (1989)；Weber, Curr Top Microbiol Immunol 1991:227-235 (1995) 和 Balakirev 等 J of Virol 76:6323-6331 (2002)。

就腺病毒蛋白酶序列或位点来说，本发明并不旨在限制于上述所提供的共有序列。本发明预计使用任何一种的腺病毒蛋白酶。在一个实施例中，腺病毒蛋白酶来自与腺病毒载体基因组相同的腺病毒血清型。

转基因

本发明的载体可包括一种或多种转基因。这样，不同的遗传能力可导入靶细胞中。在一个实施例中，转基因编码可选择的标记。在另一个实施例中，转基因编码细胞毒蛋白。这些编码细胞毒蛋白的载体可用于消除研究环境中的某些细胞或实现治疗效果。例如，在某些情况下，期望通过增加细胞毒性以增强疗效。可通过结合细胞特异性复制的细胞毒素活性和

一种或多种代谢酶的表达而得以实现，所述代谢酶例如 HSV-tk、硝基还原酶、细胞色素 P450 或使细胞能代谢 5-氟胞嘧啶 (5-FC) 为化学治疗剂 5-氟尿嘧啶 (5-FU) 的胞嘧啶脱氨酶 (CD)、羧酸酯酶 (CA)、脱氧胞苷激酶 (dCK)、嘌呤核苷磷酸化酶 (PNP)、羧肽酶 G2 (CPG2; Niculescu-Duvaz 等 J Med Chem. 2004 年 5 月 6 日;47 (10) :2651-2658)、胸苷磷酸化酶 (TP)、胸苷激酶 (TK) 或黄嘌呤鸟嘌呤磷酸核糖转移酶 (XGPRT)。这种类型的转基因还可用于赋予旁观者效应。

可导入本发明载体的额外的转基因包括能启动细胞凋亡的因子，具有作用于编码诸如结构蛋白、转录因子、聚合酶等的细胞或病原体增殖必需蛋白的 mRNAs 能力的反义分子或核酶，病原体在细胞内增殖时的病毒或其它病原体蛋白，细胞毒素蛋白，如白喉毒素、蓖麻毒素、相思豆毒素等的链，编码改造的核酸酶 (如 RNase A) 或蛋白酶 (如胰蛋白酶、木瓜蛋白酶、蛋白酶 K、羧肽酶等) 胞质变体的基因，诸如 MCP3 α 或 MIP-1 的趋化因子，来源于病毒、细菌或哺乳动物细胞的生孔蛋白，融合基因，化学疗法敏感基因和放射敏感基因。其它感兴趣的基因包括细胞因子、抗原、跨膜蛋白等，例如 IL-1、IL-2、IL-4、IL-5、IL-6、IL-10、IL-12、IL-18 或 flt3, GM-CSF, G-CSF, M-CSF, IFN- α 、 $-\beta$ 、 $-\gamma$, TNF- α 、 $-\beta$, TGF- α 、 $-\beta$, NGF, MDA-7 (黑素瘤分化相关基因-7, mda-7/白介素-24) 等。其它的例子包括，诸如 Fas、Bax、胱天蛋白酶 (Caspase)、TRAIL、Fas 配体、一氧化氮合酶 (NOS) 等的前凋亡基因；能引起细胞融合或利于细胞融合的融合基因，例如 V22、VSV 等；诸如 p53、RB、p16、p17、W9 等的肿瘤抑制基因；与细胞周期有关的基因以及编码诸如的内皮他丁 (endostatin)、血管他丁 (angiostatin) 等的抗血管生成蛋白的基因。

其它可能用于特定的遗传修饰包括 T 细胞，例如肿瘤浸润淋巴细胞 (TILs)，在此 TILs 可被修饰以增加扩张、增强细胞毒性、减少对增殖抑制剂的响应、增加淋巴因子表达等。还希望通过提供特定的表面膜蛋白如 B7、SV40 T 抗原突变体等的表达增强靶细胞易损性。

尽管任何适当的基因或编码序列都可用于本发明实施，但是某些基因或其片段是特别适用的。例如，编码在本发明实施中有用的免疫原性多

肽、毒素、免疫毒素和细胞因子的编码区。这些编码区包括上文中的那些以及包括编码如下的额外的编码区：刺激与诸如 B7、CD28、I 类 MHC、II 类 MHC、TAPs 的免疫细胞相互作用的蛋白、诸如来自 MART-I 的免疫原性序列的肿瘤相关抗原、gp100 (pmel-17)、酪氨酸酶、酪氨酸酶相关蛋白 1、酪氨酸酶相关蛋白 2、促黑素细胞激素受体、MAGE1、MAGE2、MAGE3、MAGE12、BAGE、GAGE、NY-ESO-1、-连环蛋白、MUM-1、CDK-4、胱天蛋白酶 8、KIA 0205、HLA-A2R1701、 α -胎蛋白、端粒酶催化蛋白、G-250、MUC-1、癌胚蛋白、p53、Her2/neu、磷酸丙糖异构酶、CDC-27、LDLR-FUT、端粒酶逆转录酶、PSMA、阻断抑制信号 (CTLA4 阻断) 的抗体的 cDNAs、趋化因子 (MIP1、MIP3、CCR7 配体和网硬蛋白)、抗血管发生基因包括但不限于编码 METH-1、METH-2、TrpRS 片段基因、多育曲菌素相关蛋白、催乳素片段、PEDF、vasostatin、胞外基质蛋白和生长因子/细胞因子抑制剂的不同片段、包括但不限于血管他丁、内皮他丁、kininostatin、纤维蛋白素原-E 片段、凝血酶敏感素、tumstatin、canstatin、网状内皮系统刺激素的胞外基质蛋白的不同片段、生长因子/细胞因子抑制剂包括但不限于 VEGF/VEGFR 拮抗剂、sFlt-1、sFlk、sNRP1、鼠 Flt3 配体 (mFLT3L)、血管生成素/tie 拮抗剂、sTie-2、趋化因子 (IP-10、PF-4、Gro- β 、IFN- γ (Mig)、IFN、FGF/FGFR 拮抗剂 (sFGFR)、肝配蛋白/Eph 拮抗剂 (sEphB4 和 sephrinB2)、PDGF、TGF 和 IGF-1。在本发明实施中适用的基因可编码酶类 (例如脲酶、肾素、凝血酶、金属蛋白酶、一氧化氮合酶、超氧化物歧化酶、过氧化氢酶和其它本领域技术人员已知的酶类)、酶抑制剂 (例如 α 1-抗胰蛋白酶、抗凝血酶 III、细胞或病毒蛋白酶抑制剂、纤溶酶原激活物抑制剂-1、金属蛋白酶组织抑制剂等)、囊性纤维化跨膜通道调节因子 (CFTR) 蛋白、胰岛素、营养不良素或 I 类或 II 类主要组织相容性复合物 (MHC) 抗原。也有用的是编码能调节/调控相应基因表达的多肽，能抑制细菌、寄生虫或病毒感染或其发育的多肽 (例如抗原性多肽、抗原性表位和通过竞争抑制天然蛋白作用的反式显性蛋白变体)，细胞凋亡诱导物或抑制剂 (例如 Bax、Bcl2、BclX 和其它本领域技术人员已知的)，细胞生长抑制剂 (如 p21、p16、Rb 等)，载脂蛋白 (如 ApoAI、

ApoAIV、ApoE等), 氧自由基清除剂, 具有抗肿瘤作用的多肽, 抗体, 毒素, 免疫毒素, 标记物(如 β -半乳糖苷酶、荧光素酶等)的基因, 或任何其它为本领域所公认用于治疗或预防临床症状的感兴趣的基因。更多的转基因包括编码抑制细胞分裂或信号转导的多肽, 肿瘤抑制蛋白(例如 p53、Rb、p73), 激活宿主免疫系统的多肽, 肿瘤相关抗原(如 MUC-1, BRCA-1, 诸如 E6、E7、L1、L2 等的 HPV 早期或晚期抗原), 任选地与细胞因子结合。

已显示 TRAIL 在多种转化细胞系中诱导细胞凋亡 (Jeremias I 等, *Eur J Immunol* 1998, 28:143-152 和 Walczak H 等, *Nat Med* 1999, 5:157-163)。TRAIL 的生理学作用似乎涉及先天性和获得性免疫应答(病毒感然和转化细胞的 NK 和 T-细胞调节)。尽管 TRAIL 表达分布广泛, 但是正常的细胞似乎对 TRAIL-诱导的细胞凋亡具有抗性, 据说是由于胞内蛋白 (bcl-2、IAPs、FLIP 等) 表达之故, 其减少了凋亡信号应答。

尽管目前准确的 TRAIL 抗肿瘤特异性分子机制尚未明了, 但是选择 TRIAL 作为目的转基因, 理由在于文献中业已描述的 TRAIL 的诸多特性, 认为经腺病毒 TRAIL 的晚期表达会增强细胞致死。已描述了 E1A 能增强通过 TRAIL 的致死 (E1A 和 TRAIL: Routes 等, *J Immunol.* 2000 年 10 月 15 日;165 (8):4522-7); 报道了 E1B 19K 和 55K 蛋白减小 TRAIL 的作用 (E1B 19K 和 TRAIL: Routes 等, *J Immunol.* 2000 年 10 月 15 日;165 (8):4522-7 和抗凋亡活性: Tollefson 等, *J Virol.* 2001 Oct;75 (19):8875-87); E3: 10.4K 和 14.5K (RID) 通过在溶酶体中诱导其降解除去细胞表面的 FAS 和 TRAIL 受体 (RID、FAS 和 TRAIL: Tollefson 等, *Nature.* 1998 年 4 月 16 日;392 (6677):726-30; Shisler 等, *J Virol.* 1997 年 11 月;71 (11):8299-306; Lichtenstein 等, *J Virol.* 2002 年 11 月;76 (22):11329-42; Tollefson 等, *J Virol.* 2001; E3: 14.7K 抑制了通过 TNF、Fas 和 TRAIL 的细胞凋亡, 并结合胱天蛋白酶 8 (E3 14.7K 和 TRAIL: Chen 等, *J Biol. Chem.* 1998 年 3 月 6 日;273 (10):5815-20 和 Tollefson 等, *J Virol.* 2001) 以及 E3: 6.7K 阻止了通过 TRAIL 的细胞凋亡, 并保持 ER Ca 体内平衡 (E3 6.7K 和 TRAIL/ER Ca²⁺动态平衡: Benedict 等, *J Biol. Chem.* 2001 年 2 月 2 日;276 (5):3270-8 和腺病毒的

E3-6.7K 蛋白/内质网中定位: Wilson- Rawls 等, *Virology*. 1993 年 7 月;195 (1) :6-15)。

本发明进一步包含具有协同作用、补充作用和/或无重叠毒性的两种或多种转基因的组合以及作用的方法。总之,本发明提供了在病毒载体基因组特定区域中插入转基因编码区的方法。该方法利用已知的病毒转录元件和表达 Ad 基因的机制,因为无需额外的启动子且调节信号包含较小尺寸的 DNA 片段,减小了插入在 Ad 基因组中用于转基因表达的 DNA 序列的尺寸,提供了转基因时序调节(如感染早期对晚期;感染早期对中间期)的灵活性,还提供了调节转基因表达量的技术。例如,通过将转基因插入在通常以高水平表达的转录物中和/或可操作地连接高效剪接受体位点和转基因编码区,可表达更高数量的转基因。表达水平还受调节信号与其共有序列接近程度的影响;可进行改变以适应所需表达。

在设计本发明的腺病毒载体中,要考虑到转基因的生物活性,如在某些情况下,将转基因插入在载体中使得该转基因仅在感染晚期表达或在主要在感染晚期表达(在病毒 DNA 复制后)是有利的。例如,如本文所进一步描述的,转基因可插入在 L3 中。对于某些转基因,优选在病毒生活周期早期表达转基因。在这样的情况下,转基因可插入在任一早期区(例如 E3)或 L1 区上游。

治疗方法

将有效量的本发明的载体和药学上可接受的赋形剂作为组合物给予哺乳动物(如人),所述赋形剂可包括以下一种或多种:盐水溶液、适合的缓冲液、防腐剂、稳定剂等。本发明的载体可与诸如止吐剂的合适药剂一起给予。有效量是一种足以产生有益或期望结果包括临床疗效的量。有效量可在一次或多次给药或剂量中给予。对本发明来说,载体的有效量是一种足以减轻、改善、稳定、逆转、减缓或延迟疾病状态进展,或缓解疾病的一种或多种症状的量。所要给予的量应取决于个体情况、疾病程度、给药途径、给药剂量以及期望的结果。

一般通过位点特异性注射或静脉注射实现本发明载体的转移。载体的位点特异性注射可包括,例如注射入肿瘤,以及腹膜内、胸膜内、鞘

内、动脉内、皮下注射，皮内注射，肌内注射或局部施用。这些方法提供能容易地供仅使用载体或联合使用载体和化学治疗剂的治疗之用。本发明还涉及使用载体离体感染受试者细胞。例如，分离自哺乳动物的细胞。所分离的细胞可含有肿瘤细胞和非肿瘤细胞的混合物。用复制型病毒感染细胞，病毒在肿瘤细胞中特异性复制。因此，消除了肿瘤细胞，视所需保留的非肿瘤细胞可返回给予同一哺乳动物，或视所需给予不同的哺乳动物。

如果用作包装的腺病毒，腺病毒载体可在适当的生理学上可接受的载体中以约 10^4 至约 10^{14} 的剂量给予。如果作为多核苷酸构建体（即不包装为病毒）给予，可给予约 0.01 g 至约 1000 g 的腺病毒载体。所要给予的精确剂量取决于多种因素，包括患者年龄、体重和性别，以及被治疗的肿瘤大小和严重程度。取决于预期用途和宿主潜在的免疫应答，腺病毒载体可一次或多次给予，还可多次、同时注射。如果不期望免疫应答，可通过使用多种免疫抑制剂或使用诸如从血液中除去腺病毒抗体的免疫吸附操作（如免疫血浆分离置换）的技术减少免疫应答，以便容许重复给药，而无强的免疫应答。如果包装为诸如 HSV 的不同病毒形式，所给予的量是基于与特定病毒相关的标准知识（其易于获得自，例如出版文献）且能凭经验确定。

在一个实施例中，宿主生物体是人类患者。对于人类患者，如果治疗编码区包括在载体中，则治疗编码区可以是人源的，尽管在人中可以使用具有高同源性和在生物学上相同或等价的功能的相关接近物种的基因，但前提在于所述基因在受体中不产生不利的免疫反应。核酸序列或治疗基因的治疗有效量是一种在剂量上且对于达到期望结果所需的一段时间有效的量。该量可随不同的因素而改变，包括但不限于受试者性别、年龄、体重等。

本发明的实施例包括给予本发明的癌特异性载体和另一种抗肿瘤治疗组合的方法，所述抗肿瘤治疗可包括放射、给予具有瘤形成个体抗肿瘤或化学治疗剂等，如美国申请 20030068307 所述。癌特异性载体和化学治疗剂可同时给予或连续给予，对于连续给药而言具有不同的时间间隔。在某些实施例中，有效量的载体和有效量的至少一种抗肿瘤剂或化学治疗剂与适合的赋形剂和/或缓冲溶液组合，并通过任一在此所列或本领域已知

的方法，出于相同的解决方案同时给予。当化学治疗剂并不危及载体自身生存力和/或活性时，其是可以应用的。

在给予多于一种化学治疗剂时，可在相同的组合物中一起给予；或以任何顺序连续给予；或在不同的组合物中同时给予。如果药剂连续给予，则给药可进一步包含时间延迟。连续给药可以任何顺序，并相应地包括首先给予有效量的载体，接着给予有效量的化学治疗剂。给予癌特异性载体和化学治疗剂的间隔可在至少（或者，可选择地少于）几分钟、几小时或几天之内。连续给药还包括给予选择的抗肿瘤剂，之后给予载体。给药间隔可在至少（或者，可选择地少于）几分钟、几小时或几天之内。

上述给药方法还可包括癌特异性载体和化学治疗剂的重复剂量或疗程，尤其取决于个体响应和个体疾病的特性。在第一疗程后（即一天内）马上采用重复剂量，或在间隔几天、几周或几个月后采用，以达到和/或保持对肿瘤生长的抑制。根据上述方法，特定的疗程例如联合癌特异性载体和化学疗法可随后接有放射和癌特异性载体的联合疗程。

抗肿瘤（化学治疗）剂包括来自于每种主要类型化学疗法的那些，包括但不限于：烷化剂、生物碱、抗代谢物、抗肿瘤抗生素、亚硝脲、激素激动剂/拮抗剂和类似物、免疫调节剂、光敏剂、酶等。在某些实施例中，抗癌剂是生物碱、抗代谢物、抗生素或烷化剂。在某些实施例中，抗肿瘤剂包括例如塞替派、干扰素 α -2a 和 M-VAC 组合（甲氨喋呤-长春碱、阿霉素、环磷酰胺）。优选的抗肿瘤剂包括例如 5-氟尿嘧啶、顺铂、5-氮杂胞苷和 2,2-二氟脱氧胞嘧啶核苷。特别优选的实施例包括，但不限于 5-氟尿嘧啶、2,2-二氟脱氧胞嘧啶核苷、阿霉素、miroxantrone、丝裂霉素、氮烯咪胺、卡氮芥、长春碱、罗莫司丁、三苯氧胺、紫杉萜、紫杉醇或顺铂。化学治疗剂的具体选择取决于尤其是要被治疗的疾病的特性。这些特性包括但不限于肿瘤位置、患病期以及如果有的话个体对前述治疗的响应。

除单种抗肿瘤剂与特定的癌特异性载体联合使用之外，本发明还包括超过一种的药剂与癌特异性载体联合使用。当用于治疗瘤形成时，这些抗肿瘤剂的组合通常称为联合化疗，并且取决于个体癌症的特性，常作为

还包括外科手术和或放射的联合物理疗法的一部分。预计本发明的联合癌特异性载体/化学疗法还可用作联合物理疗法程序的一部分。

本领域众所周知多种用于给予抗肿瘤剂的转移方法，包括口服和肠胃外方法。口服给予大量抗肿瘤剂具有诸多缺点，包括低生物利用度、刺激消化道以及需要用药者记录复杂的药物组合。大多数的肠胃外抗肿瘤剂给药是经静脉内的，因为肌内和皮下注射通常导致对组织的刺激或损害。肠胃外注射的区域变化包括动脉内、膀胱内、肿瘤内、鞘内、胸膜内、腹膜内和腔内注射。

化学治疗剂的转移方法包括静脉内、肠胃外和腹膜内方法以及口服给药。静脉内方法还包括通过末端静脉转移以及包括更多的部位特异性转移，例如静脉滴注入门静脉中。其它肠胃外转移方法包括直接注射抗肿瘤剂溶液，例如皮下、腔内或肿瘤内注射。

可利用本领域已知的任何一种技术来评估特定治疗方法的有效性，包括诸如成像技术，血清肿瘤标记物分析，活组织检查，肿瘤相关症状的存在、不存在或改善的诊断方法。应当理解的是所给出的治疗方法可适当改变以获得最大疗效。

在本发明的另一方面中，提供了一种包含本发明的重组病毒载体和/或颗粒以及药学上可接受的载体的药物组合物。这样的在药学上可接受的载体中可包含有效量的本发明的癌特异性载体和/或病毒颗粒的组合物适合于以单位剂型、无菌注射用溶液或口服液剂或混悬液、无菌的非注射用溶液或口服液剂或混悬液、水包油或油包水乳状液等局部或全身给予个体。本领域熟知用于肠胃外和非肠胃外药物转移的剂型。组合物还包括冻干和/或重建形式的本发明的癌特异性载体或颗粒。可接受的药用载体是，例如盐水溶液、硫酸鱼精蛋白（Elkins-Sinn, Inc., Cherry Hill, N.J.）、水、含水缓冲液例如磷酸盐缓冲液和 Tris 缓冲液、或聚凝胺（Polybrene）（Sigma Chemical, St. Louis MO）以及磷酸盐缓冲生理盐水和蔗糖。根据包含在此的教导，认为选择合适的药用载体对于本领域技术人员是显而易见的。这些溶液是无菌的，除所需的癌特异性载体外通常不含颗粒物质。所述组合物可含有接近生理条件所必需的药学上可接受的辅助物质，例如 pH 调节和缓冲剂、毒性调节剂等，例如，醋酸钠、氯化

钠、氯化钾、氯化钙、乳酸钠等。可包括增强细胞癌特异性载体摄取的赋形剂。

提出以下实施例，以便提供给本领域技术人员如何获得和使用本发明的完整内容和描述，但并不旨在限制发明人对其发明所认定的范围，也不旨在表示以下试验涵盖全部或只有这些试验得以进行。就所使用的数值（如数量、温度等）来说，已努力保证了其的精确性，但某些试验误差和偏差应予说明。除非另有说明，份数是重量份，分子量是平均分子量，温度是摄氏度以及压力是大气压或接近大气压。

所有在该说明书中引用的出版物和专利申请在此并入作为参考，相当于每篇出版物或专利申请都特别且单独地被指示并入作为参考。

依照本发明人所建立或建议的以包含用于本发明实施的优选方法的特定实施例，本发明业已得到描述。本领域技术人员应当理解，考虑到现有公开内容，可对示范性的特定实施例进行众多的修饰和改变而不背离本发明的预定范围。例如，鉴于密码子冗余，可对下划线的 DNA 序列进行改变而不影响蛋白序列。此外，鉴于考虑到生物学功能等价性，可对蛋白结构进行改变，在性质或数量上不影响生物作用。所有这样的修饰都意在包括于优选实施例的范围中。

实验例

应当理解的是本发明的方法和组合物可以多种实施例的形式体现，但仅有少数在此披露。本领域技术人员应当明白其它实施例存在且并不背离本发明的精神。因此，所描述的实施例仅是示范性的，不应视为限制。

提供以下实施例作为例证而非限制。在以下实施例中，分支点序列加剪接受体序列是：TACTTAT GACTCGTACTATTGTTATTCATCC AG↑G (SEQ ID NO:39)。下划线的序列是分支点序列，箭头表示依照剪接规则的剪接位点位置。因为该规则决定了共有序列并不是不变的，故其它遵守规则的类似序列也可使用。可选择地，可使用存在于另一种 Ad 血清型中的分支点加剪接受体位点。以下实施例涉及来源于人 Ad5 血清型的示范性的腺病毒载体。由其它腺病毒可制备类似的构建体。

实验例 1: 构建 Ad 载体并增殖编码转基因的 Ad 载体病毒粒子

以下实施例描述了在载体中插入至少一个转基因，所述载体含有腺病毒载体基因组一部分或完整的腺病毒载体基因组以及所插入的转基因。就编码完整的腺病毒载体基因组以及所插入的转基因的载体来说，将载体转染入腺病毒生产细胞系，并使用标准技术增殖病毒。在转染前，可用限制性内切酶消化载体，在病毒载体基因组内其并不切割，但切割载体（如质粒）骨架。

可选择地，转基因插入在含有部分病毒基因组的载体。在该情况下，含有部分病毒基因组的载体的一部分被克隆入含有病毒载体基因组剩余部分的载体中。因此构建了编码完整的病毒载体的载体。接着，将完整的病毒载体基因组转染进入生产细胞系中，病毒被增殖。在另一种方法中，将含有转基因和病毒基因组一部分的载体与病毒载体基因组的适当片段一起转染进入生产细胞系中，由于同源重组，会形成完整的病毒载体基因组，其将在生产细胞系中复制并增殖。

以下参考文献提供了病毒载体构建和病毒增殖的更多细节： Ghosh-Choudhury 等 *Gene*. 1986;50 (1-3) :161-71； Toietta 等 *Mol Ther*. 2002 年 2 月;5 (2) :204-10，在此并入作为参考。该实施例描述了在 Ad 的 L2 区中不同位置处插入转基因。

实验例 2: 克隆带有修饰的 Ad 载体并用于克隆在 L2 区中的进一步修饰

对于使用细菌系统的重组，有利之处在于具有在将被改变的区中也含有至少一个独特限制酶切位点的全长 Ad 基因组质粒（具有任意特定修饰的）。在以下一些实施例中，靶区是 L2 区。然而，并不存在独特位点。因此，通过以下方法构建具有这些特征的质粒。将相应于野生型 Ad 的碱基 15672-21562 的 Ad 的 BamHI 至 AscI 区插入克隆质粒 pNEB193（New England Biolabs）中，并将该质粒称为 CP 1563。进行定向诱变将独特的 SwaI 位点掺入位置 16530（相应于 Ad 野生型的碱基）中，并将该质粒称为 CP 1564。将 BamHI 至 PmeI 片段插入在克隆质粒 pMCS5 中，制备含有 Ad5 序列 13259-21562 的质粒，并称为 cp 1165。改变 cp 1165

使得其不再含有克隆载体中所存在的 *AscI* 位点，并成为 CP 1166。通过使用 *AscI* 和 *BamHI* 消化切除来自 CP 1564 的含有 *SwaI* 的 Ad 片段，并插入用 *AscI* 和 *BamHI* 切割的 CP 1566 中，并将该质粒称为 CP 1567。这就反向构建了一种全长 Ad 基因组质粒，以产生在 L2 区中含有独特 *SwaI* 位点的基因组质粒，可被切割用于 L2 区中重组。

实验例 3: 转基因插入在 L2 区中 pVII CDS 和 pV CDS 之间

在含有完整的腺病毒载体基因组（即 ITR 至 ITR）或腺病毒基因组 L2 区或其相关部分的载体（如质粒）中执行以下克隆步骤。将转基因 CDS 即开放阅读框（ORF）或 cDNA 插入在 pVII CDS 下游和 pV CDS 上游。在这两个基因之间已存在剪接受体序列，其用于表达 pV；有关的分支点序列没有精确作图。将转基因 CDS 或 cDNA 插入在内源 pV 剪接位点上游，并将额外的分支点和剪接受体位点插入在转基因上游。这些额外的剪接序列用于转基因表达。用于 L2 区的示范性的剪接受体序列不同于 L3 区所使用的。对于 pVII 根据已有的剪接位点选择新的剪接受体位点：TACTTATAGTGAA AACGTTCTGCTCTCACAG 通过保存强碱基（strong bases）以获得 TACTTATAGTAATCTAATTCCTGCTCTCTCAG（SEQ ID NO:42）。可选择地，将可操作地连接转基因 CDS 的额外的分支点和剪接受体位点插入在天然 pV 分支点和剪接受体位点上游。

实验例 4: 转基因插入在 L2 区中五邻体 CDS 和 pVII CDS 之间

将转基因 CDS 插入在五邻 CDS 下游和 pVII CDS 上游。对于 pVII 推测的分支点和剪接受体位点存在于五邻体编码区中。因此，所插入的转基因使用了存在于五邻体 CDS 中的内源剪接信号，并将额外的分支点和剪接受体位点置于转基因 CDS 之后和（并可操作地连接）pVII CDS 上游以合成 pVII 基因的 mRNA。

实验例 5: 转基因插入在 L2 区中 Mu CDS 和 L4 polyA 信号之间

将转基因插入在 Mu (aka pX) CDS 下游和 L2 多腺苷酸化信号序列上游。将分支点、剪接受体位点和转基因 CDS 插入在 Mu CDS 下游和 L2

多腺苷酸化信号序列上游。所插入的分支点和剪接受体位点可操作地连接转基因 CDS。

实验例 6: 转基因插入在 L2 区中 Mu CDS 和 L4 polyA 信号下游

将可操作地连接转基因的 IRES 插入在 Mu CDS 下游和 L2 多腺苷酸化信号序列上游。

实验例 7: 剪接元件和 hIL-24 克隆入 L3 区的 23K CDS 和 L3 polyA 之间

用 EcoRI 切割 pCR2.1 Topo (Invitrogen, Carlsbad, California), 并再连接以除去一个 EcoRI 位点。然后用 BspHI 切割该质粒, 用 Klenow 补平, 并再连接以除去 BspHI 位点。接着用 ApaI 和 RsrII 切割该质粒, 用 Klenow 补平, 并再连接以除去在这些位点之间的额外序列。保留含有氨苄青霉素抗性基因的部分。将该质粒称为 CP 1601。用 HindIII 和 XhoI 切割 pBHGE3 (Microbix Biosystems, Toronto, Ontario, Canada), 并将含有野生型 Ad5 基因组核苷酸 18318-24795 的 HindIII/XhoI 片段连接入 HindIII-和 XhoI-切割的 CP 1601 片段 (载体骨架) 中。所得到的质粒称为 CP 1603。

为在 23K CDS 下游和 L3 多腺苷酸化信号序列上游插入转基因人-IL24 (hIL-24, 也被称为 MDA-7), 并容许通过剪接机制表达, 执行以下步骤: 使用 PCR 重叠延伸拼接 (Horton 等 *Biotechniques*. 1990 年 5 月;8 (5):528-35) 含有分支点和剪接受体位点的片段, 扩增了 hIL-24 CDS。

(该 IL-24 CDS 获得自 Invivogen 质粒参见“porf-hil24”)。在 PCR 引物中编码了分支点和剪接受体位点。使用上述相同的分支点和剪接受体位点序列。PCR 产物被 TA 克隆入 pCR2.1 Topo (Invitrogen, Carlsbad, California) 构建 CP1602。然后, 将来自含有 Ad 和 hIL-24 序列的 CP1602 的 HindIII/XhoI 区连接入 HindIII/XhoI 消化的 CP 1601, 产生质粒 CP1606 Rev (SEQ ID NO:48)。将 CP1606 Rev 的 KpnI/SfiI 片段连接入 KpnI/SfiI 切割的 CP 1524 以构建 CP1610。CP1610 是用于将修饰的 L3 区“穿梭”入完整的腺病毒载体基因组中的穿梭质粒。就该质粒, 在用 BspHI 和 NheI

消化转基因片段（通常是通过 PCR 产生的，具有适当的末端）和 CP1610 后，任何转基因都可插入这里。通过同源重组 CP 1606 Rev (SEQ ID NO:48) 和腺病毒载体基因组骨架质粒，得以产生 Ad 载体或病毒。腺病毒载体基因组骨架含有所有其它必需的腺病毒基因以及提供与所要插入区同源性的 DNA 区，由此产生含有包括 L3 区修饰的腺病毒载体基因组完整序列的载体。重组科在大肠杆菌菌株中进行以产生含有带有期望改变的全长 Ad 基因组的质粒。然后，通过转染入哺乳动物细胞和病毒扩增，将该载体用于构建腺病毒病毒颗粒。

使用相同策略，其它转基因可正好插入在 L3 区的 L2 多腺苷酸化位点上游。可选择地，能用分别用其 5'和 3'末端上的限制酶切位点 BspHI 和 NheI 扩增转基因，并插入 CP1606 Rev (SEQ ID NO:48)，使用该相同策略可产生相应的 Ad 病毒。

实验例 8：克隆 mIL21 入 L3 区六邻体 CDS 和 23K CDS 之间

利用供表达的剪接将鼠白细胞介素-21 转基因插入在 L3 六邻体 CDS 下游和 23K CDS 上游，进行如下。使用本领域技术人员的常用方法，将剪接受体位点和 mIL21 cDNA (Invivogen 参见: porf-mIL21) 插入在六邻体 CDS 下游和 23K CDS 上游。在 PCR 引物中编码分支点和剪接受体位点。使用上述相同的分支点和剪接受体。在该情况下，因为对于六邻体正常的 23K 剪接信号存在于 ORF 中，故添加分支点加剪接受体位点到转基因的 3'末端，同样地，转基因可插入在六邻体上游。该策略利用了转基因的内源（天然病毒）剪接序列，并添加要被用于表达 23K CDS 的额外的分支点加剪接受体。用消化适当的限制性内切酶含有 Ad5 基因组 L3 区片段（用限制性内切酶修饰克隆和/或剪接加工信号的位点）的 PCR 产物和 hIL24 cDNA。然后将片段连接入 CP1601 等价载体中以构建 CP1623。接着，将 CP1623 的 BamHI/KpnI 片段连接入 BamHI/KpnI 切割的 CP1524（实施例 7）中以构建 CP1631，即一种穿梭载体，其中在六邻体终止密码子和 23K 起始密码子后插入改造的盒（在 hIL24 cDNA 之前的 3'剪接受体位点）。CP1623 是能用于其它以该方式要被表达的转基因的穿梭质粒。通过同源重组 CP1631 和含有期望的腺病毒基因组骨架的期望的全长

载体，获得包含插入的 mIL21 转基因的完整的腺病毒载体。然后，将所得到的质粒 CP1631 用于构建新的 Ad 病毒 OV1153，其将表达转基因。

实验例 9: 克隆 IRES 和 hIL-24 入 L3 区 23K CDS 和 L3 polyA 信号之间

使用供表达的 IRES，可将诸如 hIL-24（Invivogen 参见：porf-hil24）的转基因插入在 L3 区，正好位于 L3 多腺苷酸化位点上游和所有的 L3 编码区下游。IRES 和转基因应该包括于并翻译自所有的 L3 mRNAs。使用 PCR 重叠延伸拼接法（Horton 等 *Biotechniques*. 1990 年 5 月;8（5）：528-35），产生 PCR 产物使得 IRES 例如获得自 FMDV 或 ECMV 的可操作地连接 hIL-24 CDS，并插入在 23K CDS 下游和 L3 多腺苷酸化信号序列上游。PCR 产物被 TA 克隆入 pCR2.1 构建 CP 1604。将来自 CP 1604 含有 Ad 和 hIL-24 序列的 HindIII/XhoI 区连接入 HindIII/XhoI 消化的 CP1601-NotI 中，得到 CP1607。通过用 NotI 切割 CP1601 制备 CP1601-NotI，用 Klenow 补平末端，然后再连接质粒，使得 NotI 位点被破坏。将 CP1607 的 BamHI/SfiI 片段连接入 BamHI/SfiI 切割的 CP1603 以构建质粒 CP1609。CP1607 是能用于其它以该方式要被表达的转基因的穿梭质粒。通过同源重组 CP1609 和期望的全长腺病毒基因组骨架质粒，获得包含插入的 mFLT3L 转基因的完整的腺病毒载体。然后，通过转染入哺乳动物细胞中，将所得到的质粒用于产生 Ad 病毒。

实验例 10.构建 OV 1165 对照病毒

OV 1165 是一种包含可操作地连接 E1A 的 E2F-1 启动子的对照病毒。该载体在天然位置中携带有包装信号，并在 E2F-1 启动子上游携带多腺苷酸化信号以抑制自 LITR 的转录连读。这些包含于 OV1165 中的序列来源于质粒 pFLAr21pAE2Ff。

全长重组腺病毒基因组（pFLAr21pAE2Ff）产生如下。首先，通过在大肠杆菌中连接含有经 PCR 导入的 I-SceI 限制酶切位点的 SmaI-线性化的 pAd5LtRtSmaI 穿梭质粒和 Ad5 基因组 DNA 产生全长质粒 pFLAd5，得到含有边界为 I-SceI 位点的 Ad5 基因组的 pFLAd5。其次，用 SmaI 消化

pFLAd5, 并通过凝胶纯化含有 Ad5 左末端和右末端的片段, 自我连接以产生 pAd5LtRtSmaI。用 SmaI 消化含有 Ad5 左末端 (1006 bp) 和右末端 (582 bp) 限制性酶切片段的质粒 pAd5LtRtSmaI, 并在大肠杆菌中与 Ar21pAE2f 的基因组病毒 DNA 连接以产生全长质粒 pFLAr21pAE2Ff。

称为 CP 1585 的含有上述类型全长腺病毒衍生的基因组的质粒产生如下:

自 pFLAr21pAE2fF 质粒 (pFLAr21 Sma 穿梭质粒) 分离 4648 bp SmaI 片段, 其在 E1A 上游含有 E2F 启动子。通过 Bam HI 消化、klenow 聚合酶补平以破坏该位点并再连接, 从 pFLAr21 Sma 穿梭质粒上除去了 BamHI 位点。分离称为 pE2f LtRT 穿梭质粒的 4646 bp 质粒产物。在 BJ 5183 细胞中通过使用含有野生型 Ad5 基因组的 CV802 病毒 DNA 和线性化的 pE2f LtRt 穿梭质粒, 进行细菌同源重组产生质粒 CP 1585 (38998 bp)。最终产物 CP 1585 具有全长重组腺病毒基因组, 带有可操作地连接 E1A 的 E2F-1 启动子。CP 1585 还具有接近 L4 区的独特的 BamHI 位点 (图上位置 21796), 可用于促进基因插入在晚期区。

使用来源于 CP1585 的 36,246 个碱基对的 I-SceI 片段, 经脂质转染胺转染 A549 细胞产生 OV 1165 病毒。在 A549 细胞上培养 OV 1165 并放大到 5 个滚瓶。

通过 CsCl 离心纯化病毒, 并在 ARCA 缓冲液中配制, 产量 4.97 ml, 颗粒滴度 2.33×10^{12} vp/ml (每滚瓶 2.3×10^{12} vp)。

实验例 11. 产生 L3 载体

利用可变剪接机制, 用于产生来自 Ad5 L3 区的表达转基因的载体的穿梭质粒构建如下:

使用 BJ5183 细菌重组系统产生含有重组的、全长腺病毒基因组的质粒。

使用 BJ5183 细菌重组系统产生含有重组的、全长腺病毒基因组的质粒。在该系统的应用中, 对含有腺病毒基因组较小片段的 CP1606 和 CP 1524 穿梭质粒进行修饰。

还通过 BJ5183 细菌重组系统产生 L3 载体。将质粒骨架中含有全长腺病毒基因组的大的、线性化载体与来源于同样的腺病毒基因组的 DNA 较小片段共转化入 BJ5183 细菌具有功能性 recB、recC、sbcB 和 sbcC 基因的菌株。较小片段含有与侧接 TRAIL 编码序列的较大载体相同的值得注意的核苷酸序列延伸。大的、线性化载体与 DNA 较小片段共转化入 BJ5183 细胞，用于促进较小片段同源重组入大的、线性化载体。所得到的子代质粒含有掺入 TRAIL 编码序列 (cDNA; SEQ ID NO: 46) 的全长腺病毒衍生的基因组。

从最终的穿梭质粒 (CP 1581, 来源于 CP 1524) 切下片段，并与含有全长重组腺病毒基因组的线性化质粒 (pFLAr21pAE2Ff) 共转化。子代质粒是 pFLAr21pAE2Ff 的衍生物。重组之前，在较小的基础载体 CP 1606 Rev (SEQ ID NO:48) 中，将预有准备的改变导入 CP 1524 所含的 Ad 序列。CP1524 (SEQ ID NO:47) 和 CP1606 Rev 进一步描述如下。

CP1524 (SEQ ID NO:47) 是一种含有 pcDNA3.1+ (Invitrogen) 衍生的质粒骨架中野生型 Ad5 基因组核苷酸 18318-24795 的基础穿梭载体。Ad5 基因组核苷酸 18318-24795 的延伸含有野生型 Ad5 L3 区的一部分。

使用本领域技术人员常用的分子克隆方法构建 CP 1524。最初，用限制性内切核酸酶 BstZ17 I 和 Xba I (New England Biolabs) 消化 pcDNA3.1+。用 DNA 聚合酶 I (Klenow) (New England Biolabs) 处理所得到的消化载体，补平 Xba I 消化形成的 DNA 单链延伸。所得到的“平”末端连接到一起，形成修饰的 pcDNA3.1+ 衍生载体 CP1523，除去了额外的序列。CP 1524 的 Ad5 序列来源于 pBHGE3 (Microbix)，除在 E1 区有大的缺失之外，其是一种含有野生型 Ad5 基因组的市售质粒。用限制性核酸内切酶 HindIII 和 XhoI (New England Biolabs) 从 pBHGE3 切下代表野生型 Ad5 序列核苷酸 18318-24795 的区，连接入同样切割的 CP 1523 以产生子代质粒 CP 1524。

实验例 12. 克隆剪接元件入 L3 区 23K CDS 和 L3 polyA 之间 (CP 1606 Rev 构建)

CP 1606 Rev 含有在截短的 pCR2.1 Topo 质粒骨架 (Invitrogen) 中代表野生型 Ad5 基因组核苷酸 22181-23008 的野生型 Ad5 L3 区的修饰区段。对该区的修饰包括在 hIL24 cDNA 之前的 3'剪接受体位点 (以下简称 3' SAS)。hIL24 cDNA 侧接 BspHI (与 NcoI 相容) 和 NheI 位点, 用于置换为不同的 cDNA 序列, 如 TRAIL。

通过标准的分子克隆方法构建 CP 1606 Rev。最初, 产生包含野生型 Ad5 基因组核苷酸 22181-23008 的 PCR 片段, 对其修饰以在野生型 Ad5 23K 基因 (以下简称 23K) 终止密码子和 L3 polyA 之间包括 3' SAS 和 hIL24 cDNA。通过 PCR 剪接重叠延伸构建该片段。

简言之, 产生 3 个前体 PCR 片段。片段 PCRp 1618.95.1/2 含有用引物 1618.95.1 (SEQ ID NO:52) 和 1618.95.2 (SEQ ID NO: 56) 扩增自 pBHGE3 的野生型 Ad5 基因组核苷酸 22181-22354 的一部分。PCRp 1618.95.1/2 在其下游末端掺入 3' SAS (SEQ ID NO: 45)。

片段 PCRp 1618.97.1 (SEQ ID NO:54) /1618.97.2 (SEQ ID NO:55) 含有用引物 1618.97.1 和 1618.97.2 扩增自 pORF9-hIL24 (Invivogen) 的 hIL24 cDNA。PCRp 1618.97.1/2 分别在其上游和下游掺入 3' SAS 和野生型 Ad5 L3 区的一部分。

用引物 1618.95.5 (SEQ ID NO:56) 和 1618.95.6 (SEQ ID NO:53) 扩增野生型 Ad5 基因组核苷酸 22355-23008。片段 PCRp 1618.95.1/2 和 PCRp 1618.97.1/2 具有重叠末端。将这两个片段的混合物置于具有侧翼引物 1618.95.1 (SEQ ID NO:52) 和 1618.97.2 (SEQ ID NO:55) 的 PCR 反应中。所得到的产物 PCRp 1618.95.1/1618.97.2 使野生型 Ad5 基因组核苷酸 22181-22354 的一部分与 3' SAS 和 hIL24 cDNA 连接。

同样地, 片段 PCRp 1618.97.1/2 和 PCRp 1618.95.5/6 具有重叠末端。将这两个片段的混合物置于具有侧翼引物 1618.97.1 和 1618.95.6 的 PCR 反应中。所得到的产物 PCRp 1618.97.1/1618.95.6 具有 3' SAS 和 hIL24 cDNA 以及野生型 Ad5 基因组 22355-23008 位的一部分。

因此, PCRp 1618.95.1/1618.97.2 和 PCRp 1618.97.1/1618.95.6 在改造的 3' SAS 和 hIL24 cDNA 重叠。将这两个片段的混合物置于具有侧翼引物 1618.95.1 (SEQ ID NO:52) 和 1618.95.6 (SEQ ID NO:53) 的 PCR 反应

中。所得到的产物 PCRp 1618.95.1/6 代表了在原始位置 22354 和 22355 之间（以便导入元件位于 23K 终止密码子和 L3 polyA 位点之间）掺入改造的 3' SAS 和 hIL24 cDNA 的原始野生型 Ad5 基因组核苷酸 22181-23008。通过 Topo TA 克隆，将 PCRp 1618.95.1/6 置于载体 pCR2.1 Topo 中，产生质粒 CP1602。

为了产生用于在上述修饰的 23K 区中除去和插入转基因的基础穿梭载体，通过修饰 pCR2.1 Topo 构建质粒骨架。最初，用限制性核酸内切酶 EcoRI (New England Biolabs) 消化 pCR 2.1 Topo，并再连接以去除一个 EcoRI 位点和最初两个位点之间额外的核苷酸序列，产生 pCR2.1 TopoEcoRI。为了便于在下游克隆转基因，通过用限制性核酸内切酶 BspHI (New England Biolabs) 消化从 pCR2.1 TopoEcoRI 上除去了 BspHI 限制酶切位点，用 DNA 聚合酶 I (Klenow) 处理，然后连接被处理的载体以产生 pCR2.1 Topo.EcoRI-BspHI。为了减轻下游克隆工作量，通过用限制性核酸内切酶 ApaI 和 RsrII (New England Biolabs) 消化从 pCR.21 Topo.EcoRI-BspHI 上除去额外的序列，用 DNA 聚合酶 I (Klenow) 处理，然后连接被处理的载体以产生 CP 1601。

通过切下 CP 1602 的 HindIII 至 XhoI 片段，并将其置于同样切割的 CP 1601 载体中，构建最终的基础穿梭载体 CP1606 Rev (SEQ ID NO:48)。

为产生适于与线性化载体 pFLAr21pAE2Ff 重组的较大的穿梭质粒，从 CP1606 Rev 上切下 KpnI 至 SfiI (New England Biolabs) 片段，并置于同样切割的 CP1524 载体中，产生质粒 CP1610。

实验例 13. 产生在 L3 区 23K CDS 和 L3 polyA 之间具有剪接受体 (SA) 位点和 TRAIL 的溶瘤载体 (OV 1160)

为了产生含有改造的 3' SAS 和 TRAIL cDNA (Invivogen; 参见 pORF-TRAIL) 的 pFLAr21pAE2Ff 的重组衍生物，进行两步亚克隆。通过 PCR 自 pORF-hTTRAIL 扩增 TRAIL cDNA (SEQ ID NO: 46) 构建质粒 CP 1580。使用 NcoI-NheI 位点对该片段进行限制酶切消化，并连接入 BspHI-NheI 消化后的 CP 1606 Rev。BspHI 位点与 NcoI 位点相容。由 CP 1580 的

KpnI-SfiI 片段构建 CP1581 质粒，其含有 23K 蛋白剪接位点 3' 的 TRAIL，连接入 CP 1524 穿梭质粒用于重组。质粒 CP 1582 获得自下列 CP1581 的 XhoI-HindIII 片段与全长 pFLar21paE2f 的重组。分离含有改造的 3'SAS (SEQ ID NO:45) 和 TRAIL cDNA (SEQ ID NO:46) 的 pFLAR21pAE2Ff 的重组衍生物，并命名为 CP 1582。

通过使用 I-SceI 酶的 CP 1582 限制酶切消化，释放全长腺病毒序列 (含有新的剪接受体位点和 TRAIL)，产生 OV 1160 病毒 (SEQ ID NO:51)。

实验例 14. 产生在 L3 区 23K CDS 和 L3 polyA 之间具有 IRES 和 TRAIL 的溶瘤载体 (OV 1164)

OV1164 (SEQ ID NO:50) 在六邻体编码序列终止密码子后具有 IRES。涉及 OV1164 产生的质粒是 CP 1590、CP 1591 和 CP1592 (全长)。

在三步亚克隆后产生全长 CP1592 质粒。通过 PCR 重叠延伸拼接装配含有 HexonFmdvTrail 的质粒。为了通过重组产生具有足以产生下一个质粒的重叠的片段，在该序列两侧包含来自 Ad5 的 1kb 序列。PCR 重叠延伸拼接进行如下：

(1) 使用 1706.83.1 (SEQ ID NO:57) /1706.83.2 (SEQ ID NO:58) 扩增来自 pBHG3 的 Ad5 序列；(2) 使用寡核苷酸 1706.95.1 (SEQ ID NO:59) /1618.116.3 (SEQ ID NO:60) 扩增来自 pmcsHexFmdvTRAIL 的 HexFmdvTRAIL 片段；和 (3) 使用 1706.83.1 和 1618.116.3 装配片段，克隆新片段入 PCRstopo。

(2) CP1590 是一种含有 PCR 装配片段 (ad5HexFmdvTRAILAd5) 的 PCR2.1 topo 载体。通过重组 CP1590 的 HpaI-XhoI 片段 (Ad5HexFmdvTRAILAd5) 和用 BamHI 线性化的 CP1524 产生 CP1591。通过重组 CP1591 的 XhoI-HindIII 片段和用 SgfI 线性化的 pFLar21paE2f 产生最终全长 CP1592 质粒。

使用 Expand 聚合酶，通过寡核苷酸 1619.144.1 (SEQ ID NO:61) /1619.144.3 (SEQ ID NO:62) 从 pORF-hTRAIL 扩增 TRAIL，将 TRAIL

克隆入 pmcs HexFMDV。BSTZ171 消化扩增片段，连接入 CP1557 (pmcsHexFmdv bamHI-EcoRI) 以产生 pmcsHexFmdvTrail。

用 I-SceI 酶限制酶切消化 CP1592，释放全长腺病毒序列（含有 IRES 和 TRAIL），产生病毒。

实验例 15. 病毒生产、Trail 表达和体外生物活性

从平板上转移 A549 细胞 (1-2 X 10⁷ 细胞) 到滚瓶 (带有通气孔盖的 Falcon 35-3069 折叠表面滚瓶) 中，在第 1 天于具有 2mM L-谷氨酰胺和 10% FBS 的 RPMI 1640 培养基中。在第 4 天，用 5-10 个病毒颗粒/细胞感染细胞 3-6 小时。在感染后 72-96 小时收获细胞并贮藏在 -80°C，直至 CsCl 离心纯化。对于 OV 1160 批次 1；OV 1160 批次 2 和 OV 1164 分别获得 2.55 e12、2.05 e12 和 2.11 e12 的滴度 (通过 HPLC 测定)。通过 OD 在 260nm 处测定滴度 (病毒在含有 0.1% SDS 的 TE 中以 1:10 和 1:20 稀释，在 56°C 温育 20 分钟，在分光光度计上读出 260 和 280nm 处吸光度)。

在 A549、SW780 和 HeLa S3 细胞上，对每种病毒在 50ppc 下进行病毒生长。感染持续进行 4 小时，然后在添加 3ml 新鲜的完全培养基之前用 PBS 洗涤细胞两次。在第 3 天 (转导后 72 小时) 收获细胞，冻融 3 次并分析。表 1 显示了在第 3 天的病毒产量。

测定第 3 天 (转导后 72 小时) 的病毒产量，使用六邻体测定法滴定胶原包被的 12-孔平板上 293 细胞。表 1 显示了在第 3 天的病毒产量。

表 1.

	A549				SW780			
	总 PFU	PFU/细胞	倍数差异		总 PFU	PFU/细胞	倍数差异	
OV1165	3.50E+10	7.00E+04	1.00E+00		OV1165	4.55E+10	9.10E+04	1.00E+00
					OV1160#			
OV1160#2	2.487E+09	9.74E+03	7.19E+00		2	2.97E+09	5.94E+03	1.53E+01
					OV1160#			
OV1160#1	4.27E+09	8.54E+03	8.20E+00		1	2.98E+09	5.96E+03	1.53E+01
OV1164	1.23E+09	2.46E+03	2.85E+01		OV1164	n/a		
802	5.83E+10	1.17E+05	6.00E-01		802	8.90E+10	1.78E+05	5.11E-01
	HeLaS3							
	总 PFU	PFU/细胞	倍数差异					
OV1165	3.20E+10	6.40E+04	1.00E+00					
OV1160#2	2.45E+09	4.90E+03	1.31E+01					
OV1160#1	2.24E+09	4.48E+03	1.43E+01					
OV1164	5.50E+08	1.10E+03	5.82E+01					
802	7.30E+10	1.46E+05	4.38E-01					

六邻体测定法是一种通过免疫染色测定所给培养细胞中腺病毒六邻体蛋白产量的生物学测定法。简言之，用连续稀释的 CVL 或纯化病毒感染细胞并在 37°C 温育。感染后 48 小时，用甲醇固定细胞，然后先用抗六邻体第一抗体，接着用缀合辣根过氧化物酶（以下简称 HRP）的第二抗体检测。然后将固定的培养物暴露于二氨基联苯（以下简称 DAB）底物，通过第二抗体的 HRP 将该底物转化为在显微镜下易于观察的暗色沉淀，显影六邻体生成细胞。接着对形成暗色沉淀的斑点进行目测打分。基于每个显微镜视野的斑点打分、培养皿中视野的总数、用于感染的连续稀

释的病毒的体积以及打分斑点的稀释因子，计算每毫升的感染单位（以下简称 IU/mL）：

$$\text{IU/mL} = [(\text{每个视野的斑点打分}) \times (\text{总的视野})] / [(\text{稀释的病毒体积}) \times (\text{稀释因子})]$$

在该情况下，连续稀释的 OV1160（批次 1）、OV1160（批次 2）或 OV1164 感染 293 细胞。使用抗六邻体第一抗体（Chemicon; MAB8043）、第二抗体（Amersham Biosciences; NA931V）和 DAB 底物（Pierce; 1856090）进行该测定。补充有 1% BSA（Boehringer Mannheim; 100-350）的 PBS（Mediatech; 1-031-CV）用作为抗体稀释剂，并用于所有的中间清洗步骤。

在第 7 天，使用 MTS 测定法于多种细胞系上，测定 OV1165、2 批 OV1160、OV1164、OV802、实验室来源的野生型 Ad5 分离株（Yu 等，1999）和 Add1312（一种 E1a 缺失的复制缺陷型病毒）的 EC50。结果提供于表 2，以下是 50% 细胞致死剂量（记录为 PPC）。

表 2.

细胞/病毒	OV1165	OV1160#1	OV1160#2	OV1164	802	dl312
A549	0.12	0.07	0.21	0.86	0.02	543
肺癌						
SW780	6	2	2	52	0.63	895
转行细胞癌						
SKMel-28	133	148	230	1033	15	4787
黑色素瘤						
HT29	147	90	106	784	5	4301
直肠腺癌						
H460	0.62	0.43	0.7	10	0.02	1862
肺癌						
DLD-1	64	28	34	525	18	na
直肠腺癌						

TRAIL 表达的蛋白质印迹和 ELISA 分析

使用本领域已知的方法（如 Anton 和 Graham, *J. Virology*, 69, 4600-4606, 1995, Sambrook 和 Russell, 同上）进行蛋白质印迹分析。在 A549 细胞感染后，使用 OV1164 和 OV1160 的 CVL（粗提病毒裂解液）进行第一系列的蛋白质印迹。对于 OV1164，在 1000、100 或 10 个颗粒/细胞（ppc）下感染 A549 细胞 24、48 和 72 小时，通过 HPLC 测定的和病毒颗粒（vp）计数为 8.6×10^9 vp/ml。

还用不同体积的未滴定的 OV1160 CVL 感染 A549 细胞 72 小时。感染后，用细胞刮棒收获 A549 细胞，在 14,000 下离心 3 分钟，将来自 OV1160 或 OV1164 感染的细胞的细胞沉淀重悬于 200ul 冷的印迹裂解缓冲液中，在冰上温育 30 分钟并在使用前贮藏在 -80°C 。使用 Bradford 蛋白测定法估算每份样品中的蛋白的量。将来自 OV1160 或 OV1164 感染细胞的 10 或 30ug/ml 的总细胞蛋白与含有 100mM DTT 的溶液混合并添加染料，在 85°C 变性 5 分钟并加载到 4-12%梯度 Bis-Tris SDS-Page 凝胶（Novex from Invitrogen）上。

使用纯化的 OV1160、OV1164 和 OV1165 病毒感染 A549 细胞进行第二系列的蛋白质印迹。在 10 和 100 的 PPC 下进行 A549 细胞感染，感染后第 6、24 和 48 小时收集上清液。感染后 24 和 48 小时收集细胞。30 μg 的总蛋白或 30 μl 的上清液加载到 4-12%梯度 Bis-Tris SDS-Page 凝胶（Novex from Invitrogen）上。凝胶在 1X MOPS 缓冲液中于 100 伏 1.5 小时，接着转移到 PDVF 膜上，在 10% MeOH 的 1X MOPS 缓冲液中于 400 毫安 1 小时。使用的第一抗体是山羊 IgG 抗人 TRAIL 多克隆抗体（R&D Systems; AF375）。使用山羊抗-HRP 第二抗体（Santa Cruz Bio SC-2056）。在暴露于 Biomax MS 膜（Kodak）后使用 ECL-Plus 试剂盒（Amersham）检测结果。

在 OV1160 和 OV164 以 10、100 和 1000 PPC 感染 24、48 或 72 小时的 A549 细胞的细胞裂解液中，通过蛋白质印迹检测 TRAIL 蛋白。在 OV1160 以 100 的 PCC 感染 48 小时的 A549 细胞上清液中，还检测了 TRAIL 的 19 kDa 切割部分（可溶形式）。

利用 ELISA 定量 TRAIL 表达，通过评估细胞凋亡程度，表现病毒感染细胞产生的 TRAIL 的旁观者效应（在进一步描述如下的生物测定法中评估）。进行分析以便确定病毒感染后是否产生 TRAIL 以及其是否是生物学活性的。

分别用 OV1160、OV1164 和 OV1165 在 100 的 PPC 下感染 A549 细胞 24 和 48 小时。收集上清液并冷冻在 -80°C，直至利用 ELISA 进行检测以测定上清液中 TRAIL 的量。

收集条件培养基并用夹心 ELISA 测定 TRAIL 表达。用 0.1M 碳酸盐 pH 9.6 缓冲液中的小鼠抗人 TRAIL 单克隆抗体 (BioSource) 包被 96-孔微量滴定板，并在 4°C 温育过夜。用 PBS-0.05% Tween-20 泛洗平板，并用 PBS-1% BSA-0.05% Tween-20 缓冲液封闭 1 小时。重组人 TRAIL 蛋白 (R&D Systems, Minneapolis, MN) 在连续稀释后用来做标准曲线。在孔中温育样品 1 小时，泛洗并与山羊抗人 TRAIL 多克隆抗体 (R&D Systems, Minneapolis, MN) 温育 1 小时。泛洗后，样品与 HRP-缀合的抗山羊 IgG 抗体 (Sigma Chemical Co.) 温育 1 小时，再次洗涤，并用 Sure Blue TMB 底物 (KPL, Gaithersburg, MD) 在 450/650 nm 吸光度处检测。使用重组的人 TRAIL (R&D Systems; 375 TEC) 产生标准曲线。该测定的结果提供于以下表 3A 中。

表 3A.

	PPC100			
	24hr		平均值	TRAIL (ng/ml)
OV1160#1	0.093	0.11	0.1015	1.1
OV1160#2	0.1	0.1	0.1	0.9
OV1165	0.078	0.079	0.0785	
OV1164				

	48hr		平均值	TRAIL (ng/ml)
OV1160#1	0.192	0.191	0.1915	9.6
OV1160#2	0.208	0.207	0.2075	13.3
OV1165	0.079	0.08	0.0795	
OV1164	0.082	0.08	0.081	

对 OV1160 或 OV1165 在其相应的 EC₅₀ (在第 7 天通过 MTS 测定) 感染 40 小时的 6 种细胞系 (SW780、A549、DLD-1、SKMel28、HT29、H460) 进行第二次 ELISA。进行该研究以估算多种不同细胞系产生的 TRAIL 量。表 3B 所示结果表明 OV1160 感染的 SW780、A549 和 HT460 细胞产生了最多数量的 TRAIL。

表 3B.

TRAIL (ng/ml)	SW780	A549	DLD1	SK-Mel28	HT29	H460	
平均值	OV1165	0	0	0.201301	0	0	0
	OV1160#1	10.048	110.3783	4.790433	1.839404	0.727786	13.83983
	OV1160#2	5.921467	87.18996	5.231936	1.890064	0.616181	13.55111
sd		0	0	0.284682	0	0	0
		1.997327	7.356895	0.159734	0.365634	0.072695	0
			5.811357	1.379575	0.437278	0.04107	1.353553

通过将感染细胞上清液转移到新鲜的未感染细胞上, 评估 OV1160、OV1164 和 OV1165 感染细胞产生 TRAIL 的旁观者效应。使用 ELISA 试剂盒 (细胞死亡检测试剂盒 Elisa plus/Roche 1774425) 通过 DNA 断裂评估细胞凋亡程度。

用 OV1165、OV 1164 或两批 OV1160 中的一批于 100 PPC 感染 A549 细胞 48 小时，收集感染后的 100 ul 上清液，添加或不添加常用的胱天蛋白酶抑制剂（R&D systems: FMK001），进行第一次测定。添加上清液 16 小时后进行分析。表 4 所示的结果表明对于 SW780 细胞，是 OV1160 而非 OVI 165 能诱导细胞凋亡，当胱天蛋白酶抑制剂包含在培养物中时，效果降低。

表 4.

	SW780	A549
仅有细胞	0.35	0.24
1165	0.68	0.26
1165 + 抑制剂	1	0.3
1160	3	0.175
1160 + 抑制剂	0.56	0.15
rTrail	3.75	1.45
来自试剂盒的阳性对照	3.3	
培育缓冲液	0.21	
ABTS (dye)	0.18	

对 OV1165 或 OV1160 在其相应的 EC50s 感染的 6 种细胞系（SW780、DLD-1、HT29、SkMel28、H460 和 A549）进行第二次测定，所述 EC50s 在第 7 天通过 MTS 测定。向相应的未感染细胞添加来自每种病毒感染细胞系的新鲜收获的上清液，在进行 ELISA 前 40 小时，添加或

不添加胱天蛋白酶抑制剂。如表 3B 所示，不同细胞系产生的 TRAIL 水平不同。该研究被设计来评估特定细胞系产生的 TRAIL 的生物活性。

实验例 16. 在皮下人膀胱 SW780 异种移植肿瘤模型中的抗肿瘤功效

使用人膀胱 TCC 细胞系 SW780 进行研究，以评估表达 TRAIL 病毒（OV1160）和对照（OV1165）病毒的抗肿瘤功效。在最近的研究中，分别通过五次肿瘤内注射 OV1160 和 OV1165 治疗裸鼠中 SW780 细胞的异种移植肿瘤。所测量的参数包括存活率、肿瘤体积改变、体重以及病毒感染程度、TRAIL 表达和肿瘤中细胞凋亡程度。

通过皮下注射 2×10^6 SW780 细胞到雌性小鼠（每组 10 只动物）中建立肿瘤。当平均肿瘤体积达到大约 150 mm^3 时，开始肿瘤内病毒治疗。开始于研究日（SD 1），肿瘤注射五次，每次注射剂量为 1×10^8 、 1×10^9 或 1×10^{10} 病毒颗粒（vp），每隔一天总计给药 5 次。每周测量两次肿瘤体积 $[(W^2 \times L)/2]$ ；每周测量一次体重。在感染后的不同时间点进行六邻体染色和 TRAIL ELISA。

序列表

以下是实施本发明所使用序列的描述。

LLNFDLLKLAGDVESNPGP (SEQ ID NO: 1)

TLNFDLLKLAGDVESNPGP (SEQ ID NO:2);

LLKLAGDVESNPGP (SEQ ID NO:3)

NFDLLKLAGDVESNPGP (SEQ ID NO:4)

QLNFDLLKLAGDVESNPGP (SEQ ID NO:5)

APVKQTLNFDLLKLAGDVESNPGP (SEQ ID NO:6).

VTELLYRMKRAETYCPRLLAIHPTEARHKQKIVAPVKQTLNFDLL
KLAGDVESNPG P (SEQ ID NO:7)

LLAffIPTEARHKQKIVAPVKQTLNFDLLKLAGD VESNPGP (SEQ ID
NO:8)

EARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP (SEQ ID NO:9)

NFDLLKLAGDVESNPGPFF (SEQ ID NO: 10)

GIFNAHYAGYFADLLIHDVETNPGP (SEQ ID NO: 11)

RIFNAHYAGYFADLLIHDVETNPGP (SEQ ID NO: 12)

HVFETHYAGYFADLLIHDVETNPGP (SEQ ID NO: 13)

KAVRGYHADYYKQRLIHDVEMNPGP (SEQ ID NO: 14)

RAVRAYHADYYKQRLIHDVEMNPGP (SEQ ID NO: 15)

KAVRGYHADYYRQRLIHDVETNPGP (SEQ ID NO: 16)

LTFNFDLLKLAGDVESNPGP (SEQ ID NO: 17)

LLNFDLLKLAGDMESNPGP (SEQ ID NO: 18)

MCNFDLLKLAGDVESNPGP (SEQ ID NO: 19)

CTNYALLKLAGDVESNPGP (SEQ ID NO:20)

GATNFSLLKLAGDVELNPGP (SEQ ID NO:21)

GPGATNFSLLKQAGDVEENPGP (SEQ ID NO:22)

EAARQMLLLLSGDVETNPGP (SEQ ID NO:23)

FLRKRTQLLMSGDVESNPGP (SEQ ID NO:24)

GSWTDILLLLSGDVETNPGP (SEQ ID NO:25)

RAEGRGSLTTCGDVEENPGP (SEQ ID NO:26)

TRAEIEDELIRAGIESNPGP (SEQ ID NO:27)
SKFQIDRILISGDIELNPGP (SEQ ID NO:28)
AKFQIDKILISGDVELNPGP (SEQ ID NO:29)
SKFQIDKILISGDIELNPGP (SEQ ID NO:30)
SSIIRTKMLVSGDVEENPGP (SEQ ID NO:31)
CDAQRQKLLLSGDIEQNPGP (SEQ ID NO:32)
SEQ ID NO:33 是弗林蛋白酶共有序列或位点: RXK(R)R
SEQ ID NO:34 是凝血因子 XA 切割序列或位点: IE(D)GR
SEQ ID NO:35 是信号肽酶 I 切割序列或位点: LAGFATVAQA
SEQ ID NO:36 是凝血酶切割序列或位点: LVPRGS
SEQ ID NO:37 是腺病毒共有蛋白酶序列或位点(M,L,I)XGG/X
SEQ ID NO:38 是腺病毒共有蛋白酶序列或位点(M,L,I)XGX/G
SEQ ID NO:39 是分支点加剪接受体的示范性序列
SEQ ID NO:40 是分支点共有序列
SEQ ID NO:41 是 GenBank AY339865 的核苷酸 29209-29336
SEQ ID NO:42 是 L2 区剪接位点
SEQ ID NO:43 是 E2F 启动子序列
SEQ ID NO:44 是 FMDV (毒株 C) IRES 序列
SEQ ID NO:45 是 3'剪接受体位点或 SAS
SEQ ID NO:46 是 hTRAIL ORF (INVIVOGEN)
SEQ ID NO:47 是 CP1524 序列 (9587 个碱基对)
SEQ ID NO:48 是 CP 1606 REV 序列 (3855 个碱基对)
SEQ ID NO:49 是 OV 1165 序列
SEQ ID NO:50 是 OV1164 序列
SEQ ID NO:51 是 OV1160 序列
SEQ ID NO:52 是寡核苷酸引物 1618.95.1 的序列
SEQ ID NO:53 是寡核苷酸引物 1618.95.的序列
SEQ ID NO:54 是寡核苷酸引物 1618.97.1 的序列
SEQ ID NO:55 是寡核苷酸引物 1618.97.2 的序列
SEQ ID NO:56 是寡核苷酸引物 1618.95.5 的序列

SEQ ID NO:57 是寡核苷酸引物 1706.83.1 的序列
 SEQ ID NO:58 是寡核苷酸引物 1706.83.2 的序列
 SEQ ID NO:59 是寡核苷酸引物 1706.95.1 的序列
 SEQ ID NO:60 是寡核苷酸引物 1618.116.3 的序列
 SEQ ID NO:61 是寡核苷酸引物 1619.144.1 的序列
 SEQ ID NO:62 是寡核苷酸引物 1619.144.3 的序列

以下是实施本发明所使用序列的描述。

LLNFDLLKLAGDVESNPGP(SEQ ID NO:1)

TLNFDLLKLAGDVESNPGP(SEQ ID NO:2);

LLKLAGDVESNPGP (SEQ ID NO:3)

NFDLLKLAGDVESNPGP (SEQ ID NO:4)

QLLNFDLLKLAGDVESNPGP(SEQ ID NO:5)

APVKQTLNFDLLKLAGDVESNPGP (SEQ ID NO:6).

VTELLYRMKRAETYCPRPLLAIHPTEARHKQKIVAPVKQTLNFDLL
 KLAGDVESNPG P (SEQ ID NO:7)

LLAFFLPTEARHKQKTVAPVKQTLNFDLLKLAGDVESNPGP (SEQ
 ID NO:8)

EARHKQKJVAPVKQTLNFDLLKLAGD VESNPGP (SEQ ID NO:9)

NFDLLKLAGDVESNPGPFF (SEQ ID NO: 10)

GIFNAHYAGYFADLLIHDVETNPGP (SEQ ID NO: 11)

RIFNAHYAGYFADLLIHDVETNPGP (SEQ ID NO: 12)

HVFETHYAGYFADLLIHDVETNPGP (SEQ ID NO: 13)

KAVRGYHADYYKQRLIHDVEMNPGP (SEQ ID NO: 14)

RAVRAYHADYYKQRLIHDVEMNPGP (SEQ ID NO: 15)

KAVRGYHADYYRQRLIHDVETNPGP (SEQ ID NO: 16)

LTNFDLLKLAGDVESNPGP (SEQ ID NO: 17)

LLNFDLLKLAGDMESNPGP (SEQ ID NO: 18)

MCNFDLLKLAGDVESNPGP (SEQ ID NO: 19)

CTNYALLKLAGDVESNPGP (SEQ ID NO:20)

GATNFSLLKLAGDVELNPGP (SEQ ID NO.-21)

GPGATNFSLLKQAGDVEENPGP (SEQ ID NO:22)

EAARQMLLLLSGDVETNPGP (SEQ ID NO:23)

FLRKRTQLLMSGDVESNPGP (SEQ ID NO-.24)

GSWTDILLLLLSGDVETNPGP (SEQ ID NO:25)

RAEGRGSLTTCGDVEENPGP (SEQ ID NO:26)

TRAEIEDELIRAGIESNPGP (SEQ ID NO:27)

SKFQIDRILISGDIELNPGP (SEQ ID NO:28)

AKFQIDKILISGDVELNPGP (SEQ ID NO-.29)

SKPQIDKILISGDIELNPGP (SEQ ID NO:30)

SS[Pi]RTKMLVSGDVEENPGP (SEQ ID NO.-31)

CDAQRQKLLLSGDIEQNPGP (SEQ ID NO:32)

SEQ ID NO:33 是弗林蛋白酶共有序列或位点: RXK(R)R

SEQ ID NO:34 是凝血因子 XA 切割序列或位点: IE(D)GR

SEQ ID NO:35 是信号肽酶 I 切割序列或位点: LAGFATVAQA

SEQ ID NO:36 是凝血酶切割序列或位点: LVPRGS

SEQ ID NO:37 是腺病毒共有蛋白酶序列或位点(M,L,I)XGG/X

SEQ ID NO:38 是腺病毒共有蛋白酶序列或位点(M,L,I)XGX/G

SEQ ID NO:39 是分支点加剪接受体的示范性序列

TACTTATGACTCGTACTAT

TGTTATTCATCC AG6G

SEQ ID NO:40 是分支点共有序列 YNYURAY (在此 Y 是嘧啶, N 是任何核苷酸, 和 R 是嘌呤)

SEQ ID NO:41 是标号 GenBank AY339865 的核苷酸 29209-29336

TAATTTACTAAGTTACAAAGCTAATGTCACCACTAACTGCTTTACTCGCTGCTTG

CAAAACAAATTCAAAAAGTTAGCATTATAATTAGAATAGGATTTAAACCCCCCG

GTCATTTCTGCTCAATAC

SEQ ID NO :42 是 L2 区剪接位点

TAC TTAT AGT AAT CTA A TT CCT G CT CTC TC AG

SEQ ID NO:43 是 E2F 启动子序列

CATCCGGACAAAGCCTGCGCGCGCCCCGCCCGCCATTGGCCGTACCGCCCCCGC

GCCGCCGCCCATCTCGCCCCCTCGCCGCCGGGTCCGGCGCGTTAAAGCCAATAG
GAACCGCCGCCGTTGTTCCCGTCACGGCCGGGGCAGCCAATTGTGGCGGGCGCTC
GGCGGCTCGTGGCTCTTTCGCGGCAAAAAGGATTTGGCGCGTAAAAGTGGCCGG
GACTTTGCAGGCAGCGGGCGGCCGGGGGCGGAGCGGGATCGAGCCCTCGATGAT
ATCA

SEQ ID NO:44 是 FMDV (毒株 C) IRES 序列

AGCAGGTTTCCCAACTG
ACACAAAACGTGCAACTTGAAACTCCGCCTGGTCTTCCAGGTCTAGAGGGGTAACA
CTTTGTA CTGCGT
TTGGCTCCACGCTCGATCCACTGGCGAGTGTTAGTAACAGCACTGTTGCTTCGTAGCG
GAGCATGACGGC
CGTGGGAACTCCTCCTTGGTAACAAGGACCCACGGGGCCAAAAGCCACGCCACACG
GGCCCGTCATGTG
TGCAACCCCAGCACGGCGACTTTACTGCGAAACCCACTTTAAAGTGACATTGAAACT
GGTACCCACACAC
TGGTGACAGGCTAAGGATGCCCTTCAGGTACCCCGAGGTAACACGCGACACTCGGGA
TCTGAGAAGGGGA
CTGGGGCTTCTATAAAAAGCGCTCGGTTTAAAAAGCTTCTATGCCTGAATAGGTGACC
GGAGGTCGGCACC
TTTCCTTTGCAATTAATGACCCT

SEQ ID NO:45 是 3'剪接受体位点或 SAS

TACTTATGACTCGTACTATTGTTATTCATCCAG G

SEQ ID NO:46 是 hTRAIL ORF (INVIVOGEN) (SEQ ID NO:46)

ATGGCTATGATGGAGGTCCAGGGGGGACCCAGCCTGGGACAGACCTGCGTGCTGATC
GTGATCTTTACAGTG
CTCCTGCAGTCTCTGTGTGGCTGTA ACTTACGTGTACTTTACCAACGAGCTGAAGC
AGATGCAGGACAAG
TACTCCAAAAGTGGCATTGCTTGTCTTAAAAGAAGATGACAGTTATTGGGACCCCA
ATGACGAAGAGAGT
ATGAACAGCCCCTGCTGGCAAGTCAAGTGGCAACTCCGTCAGCTCGTTAGAAAGATG

ATTTTGAGAACCTCT
GAGGAAACCATTTCTACAGTTCAAGAAAAGCAACAAAATATTTCTCCCCTAGTGAGA
GAAAGAGGTCCTCAG
AGAGTAGCAGCTCACATAACTGGGACCAGAGGAAGAAGCAACACATTGTCTTCTCCA
AACTCCAAGAATGAA
AAGGCTCTGGGCCGCAAATAAACTCCTGGGAATCATCAAGGAGTGGGCATTCATTC
CTGAGCAACTTGAC
TTGAGGAATGGTGAAGTGGTCATCCATGAAAAAGGGTTTTACTACATCTATTCCCAA
CATACTTTCGATTT
CAGGAGGAAATAAAAGAAAACACAAAGAACGACAAACAAATGGTCCAATATATTTA
CAAATACACAAGTTAT
CCTGACCCTATATTGTTGATGAAAAGTGCTAGAAATAGTTGTTGGTCTAAAGATGCAG
AATATGGACTCTAT
TCCATCTATCAAGGGGAATATTTGAGCTTAAGGAAAATGACAGAATTTTTGTTTCTG
TAACAAATGAGCAC
TTAATAGACATGGACCATGAAGCCAGTTTTTCGGGGCCTTTTTAGTTGGCTAA
SEQ ID NO:47 是 CP1524 序列 (9587 个碱基对)
GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACCTCTCAGTACAATCTGCTCT
GATGCCGCATAGTTAAG
CCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAA
TTAAGCTACAACAAGG
CAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCT
GCTTCGCGATGTACGGG
CCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACG
GGGTCATTAGTTCATAG
CCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCCGCTGGCTGAC
CGCCCAACGACCCCCGC
CCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCA
TTGACGTCAATGGGTGG
AGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGT

ACGCCCCCTATTGACGT
CAATGACGGTAAATGGCCCCGCTGGCATTATGCCAGTACATGACCTTATGGGACT
TTCCTACTTGGCAGTAC
ATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAA
TGGGCGTGGATAGCGGT
TTGACTCACGGGGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTGT
GGCACCAAATCAACG
GGACTTTCCAAAATGTCGTAACAACCTCCGCCCCATTGACGCAAATGGGCGGTAGG
CGTGTACGGTGGGAGGTC
TATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGA
AATTAATACGACTCACT
ATAGGGAGACCCAAGCTGGCTAGCGTTTAACTTAAGCTTGATCCCCGCCCTCCCG
TAGAGGAGCCTCCACCG
GCCGTGGAGACAGTGTCTCCAGAGGGGCGTGGCGAAAAGCGTCCGCGCCCCGACA
GGGAAGAACTCTGGTGA
CGCAAATAGACGAGCCTCCCTCGTACGAGGAGGCACTAAAGCAAGGCCTGCCAC
CACCCGTCCCATCGCGCC
CATGGCTACCGGAGTGCTGGGCCAGCACACCCGTAACGCTGGACCTGCCTCCCC
CCGCCGACACCCAGCAG
AAACCTGTGCTGCCAGGCCCGACCGCCGTTGTTGTAACCCGTCCTAGCCGCGCGTC
CCTGCGCCGCGCCGCA
GCGGTCCGCGATCGTTGCGGCCCGTAGCCAGTGGCAACTGGCAAAGCACACTGAA
CAGCATCGTGGGTCTGGG
GGTGCAATCCCTGAAGCGCCGACGATGCTTCTGAATAGCTAACGTGTCGTATGTGT
GTCATGTATGCGTCCAT
GTCGCCGCCAGAGGAGCTGCTGAGCCGCCGCGCGCCCGCTTTCCAAGATGGCTAC
CCCTTCGATGATGCCGCA
GTGGTCTTACATGCACATCTCGGGCCAGGACGCCTCGGAGTACCTGAGCCCCGGGC
TGGTGCAGTTTGCCCGC
GCCACCGAGACGTACTTCAGCCTGAATAACAAGTTTAGAAACCCACGGTGGCGC

CTACGCACGACGTGACCA
CAGACCGGTCCCAGCGTTTGACGCTGCGGTTTCATCCCTGTGGACCGTGAGGATACT
GCGTACTCGTACAAGGC
GCGGTTACCCTAGCTGTGGGTGATAACCGTGTGCTGGACATGGCTTCCACGTACT
TTGACATCCGCGGCGTG
CTGGACAGGGGCCCTACTTTTAAGCCCTACTCTGGCACTGCCTACAACGCCCTGGC
TCCCAAGGGTGCCCCAA
ATCCTTGCGAATGGGATGAAGCTGCTACTGCTCTTGAAATAAACCTAGAAGAAGA
GGACGATGACAACGAAGA
CGAAGTAGACGAGCAAGCTGAGCAGCAAAAACTCACGTATTTGGGCAGGCGCCT
TATTCTGGTATAAATATT
ACAAAGGAGGGTATTCAAATAGGTGTGCAAGGTCAAACACCTAAATATGCCGATA
AAACATTTCAACCTGAAC
CTCAAATAGGAGAATCTCAGTGGTACGAACTGAAATTAATCATGCAGCTGGGAG
AGTCCTTAAAAAGACTAC
CCCAATGAAACCATGTTACGGTTCATATGCAAAACCCACAAATGAAAATGGAGGG
CAAGGCATTCTTGTAAG
CAACAAAATGGAAAGCTAGAAAGTCAAGTGGAAATGCAATTTTTCTCAACTACTG
AGGCGACCGCAGGCAATG
GTGATAACTTGACTCCTAAAGTGGTATTGTACAGTGAAGATGTAGATATAGAAACC
CCAGACACTCATATTTT
TTACATGCCCACTATTAAGGAAGGTAACCTCACGAGAACTAATGGGCCAACAACT
ATGCCCAACAGGCCTAAT
TACATTGCTTTTAGGGACAATTTTATTGGTCTAATGTATTACAACAGCACGGGTAA
TATGGGTGTTCTGGCGG
GCCAAGCATCGCAGTTGAATGCTGTTGTAGATTTGCAAGACAGAAACACAGAGCT
TTCATACCAGCTTTTGCT
TGATTCCATTGGTGATAGAACCAGGTACTTTTCTATGTGGAATCAGGCTGTTGACA
GCTATGATCCAGATGTT
AGAATTATTGAAAATCATGGAACCTGAAGATGAACTTCCAAATTAAGTCTTTCCACT

GGGAGGTGTGATTAATA
CAGAGACTCTTACCAAGGTAAAACCTAAAACAGGTCAGGAAAATGGATGGGAAA
AAGATGCTACAGAATTTTC
AGATAAAAATGAAATAAGAGTTGGAAATAATTTTGCCATGGAAATCAATCTAAAT
GCCAACCTGTGGAGAAAT
TTCCTGTACTCCAACATAGCGCTGTATTTGCCGACAAGCTAAAGTACAGTCCTTC
CAACGTAAAAATTTCTG
ATAACCCAAACACCTACGACTACATGAACAAGCGAGTGGTGGCTCCCGGGTTAGT
GGACTGCTACATTAACCT
TGGAGCACGCTGGTCCCTTGACTATATGGACAACGTCAACCCATTTAACCACCACC
GCAATGCTGGCCTGCGC
TACCGCTCAATGTTGCTGGGCAATGGTCGCTATGTGCCCTTCCACATCCAGGTGCC
TCAGAAGTTCTTTGCCA
TTAAAAACCTCCTTCTCCTGCCGGGCTCATAACCTACGAGTGGAACCTCAGGAAG
GATGTTAACATGGTTCT
GCAGAGCTCCCTAGGAAATGACCTAAGGGTTGACGGAGCCAGCATTAAAGTTTGAT
AGCATTTGCCTTTACGCC
ACCTTCTTCCCATGGCCCAACACCCGCTCCACGCTTGAGGCCATGCTTAGAAA
CGACACCAACGACCAGT
CCTTTAACGACTATCTCTCCGCCGCAACATGCTCTACCCTATAACCCGCCAACGCT
ACCAACGTGCCCATATC
CATCCCTCCCGCAACTGGGCGGCTTTCGCGGCTGGGCCTTACGCGCCTTAAGA
CTAAGGAAACCCCATCA
CTGGGCTCGGGCTACGACCCTTATTACACCTACTCTGGCTCTATACCCTACCTAGAT
GGAACCTTTTACCTCA
ACCACACCTTTAAGAAGGTGGCCATTACCTTTGACTCTTCTGTCAGCTGGCCTGGC
AATGACCGCCTGCTTAC
CCCAACGAGTTTGAATTAAGCGCTCAGTTGACGGGGAGGGTTACAACGTTGCC
CAGTGTAACATGACCAA
GACTGGTTCCTGGTACAAATGCTAGCTAACTACAACATTGGCTACCAGGGCTTCTA

TATCCCAGAGAGCTACA
AGGACCGCATGTACTCCTTCTTTAGAACTTCCAGCCCATGAGCCGTCAGGTGGTG
GATGATACTAAATACAA
GGACTACCAACAGGTGGGCATCCTACACCAACACAACAACCTCTGGATTTGTTGGCT
ACCTTGCCCCCACCATG
CGCGAAGGACAGGCCTACCCTGCTAACTTCCCCTATCCGCTTATAGGCAAGACCGC
AGTTGACAGCATTACCC
AGAAAAAGTTTCTTTGCGATCGCACCCCTTTGGCGCATCCCATTCTCCAGTAACTTTA
TGTCCATGGGCGCACT
CACAGACCTGGGCCAAAACCTTCTCTACGCCAACTCCGCCACGCGCTAGACATGA
CTTTTGAGGTGGATCCC
ATGGACGAGCCCACCCTTCTTTATGTTTTGTTTGAAGTCTTTGACGTGGTCCGTGTG
CACCGGCCGCACCGCG
GCGTCATCGAAACCGTGTACCTGCGCACGCCCTTCTCGGCCGGCAACGCCACAACA
TAAAGAAGCAAGCAACA
TCAACAACAGCTGCCGCCATGGGCTCCAGTGAGCAGGAACTGAAAGCCATTGTCA
AAGATCTTGGTTGTGGGC
CATATTTTTTGGGCACCTATGACAAGCGCTTCCAGGCTTTGTTTCTCCACACAAGC
TCGCCTGCGCCATAGT
CAATACGGCCGGTCGCGAGACTGGGGGCGTACACTGGATGGCCTTTGCCTGGAAC
CCGCACTCAAAAACATGC
TACCTCTTTGAGCCCTTTGGCTTTTCTGACCAGCGACTCAAGCAGGTTTACCAGTTT
GAGTACGAGTCACTCC
TGCGCCGTAGCGCCATTGCTTCTTCCCCGACCGCTGTATAACGCTGGAAAAGTCC
ACCCAAAGCGTACAGGG
GCCCAACTCGGCCGCCTGTGGACTATTCTGCTGCATGTTTCTCCACGCCTTTGCCAA
CTGGCCCCAAACTCCC
ATGGATCACAACCCACCATGAACCTTATTACCGGGGTACCCAACTCCATGCTCAA
CAGTCCCCAGGTACAGC
CCACCCTGCGTCGCAACCAGGAACAGCTCTACAGCTTCCCTGGAGCGCCACTCGCCC

TACTTCCGCAGCCACAG
TGCGCAGATTAGGAGCGCCACTTCTTTTTGTCACCTTGAAAAACATGTAAAAATAAT
GTACTAGAGACACTTTC
AATAAAGGCAAATGCTTTTATTTGTACACTCTCGGGTGATTATTTACCCCCACCCTT
GCCGTCTGCGCCGTTT
AAAAATCAAAGGGGTTCTGCCGCGCATCGCTATGCGCCACTGGCAGGGACACGTT
GCGATACTGGTGTTTAGT
GCTCCACTTAAACTCAGGCACAACCATCCGCGGCAGCTCGGTGAAGTTTTCACTCC
ACAGGCTGCGCACCATC
ACCAACGCGTTTAGCAGGTCGGGCGCCGATATCTTGAAGTCGCAGTTGGGGCCTCC
GCCCTGCGCGCGGAGT
TGCGATACACAGGGTTGCAGCACTGGAACACTATCAGCGCCGGGTGGTGCACGCT
GGCCAGCACGCTCTTGTC
GGAGATCAGATCCGC
GGAGATCAGATCCGCGTCCAGGTCCTCCGCGTTGCTCAGGGCGAACGGAGTCAAC
TTTGGTAGCTGCCTTCCC
AAAAAGGGCGCGTGCCAGGCTTTGAGTTGCACTCGCACCGTAGTGGCATCAAAA
GGTGACCGTGCCCGGTCT
GGGCGTTAGGATACAGCGCCTGCATAAAAGCCTTGATCTGCTTAAAAGCCACCTG
AGCCTTTGCGCCTTCAGA
GAAGAACATGCCGCAAGACTTGCCGGAAAAGTATTGGCCGGACAGGCCGCGTCG
TGCACGCAGCACCTTGCG
TCGGTGTTGGAGATCTGCACCACATTTGGCCCCACCGTTCTTCACGATCTTGGC
CTTGCTAGACTGCTCCT
TCAGCGCGCGCTGCCCGTTTTCGCTCGTCACATCCATTTCAATCACGTGCTCCTTAT
TTATCATAATGCTTCC
GTGTAGACACTTAAGCTCGCCTTCGATCTCAGCGCAGCGGTGCAGCCACAACGCGC
AGCCCGTGGGCTCGTGA
TGCTTGTAGGTACCTCTGCAAACGACTGCAGGTACGCCTGCAGGAATCGCCCCAT
CATCGTCACAAAGGTCT

TGTTGCTGGTGAAGGTCAGCTGCAACCCGCGGTGCTCCTCGTTCAGCCAGGTCCTG
CATACGGCCGCCAGAGC
TTCCACTTGGTCAGGCAGTAGTTTGAAGTTCGCCTTTAGATCGTTATCCACGTGGTA
CTTGTCATCAGCGCG
CGCGCAGCCTCCATGCCCTTCTCCCACGCAGACACGATCGGCACACTCAGCGGGTT
CATCACCGTAATTTAC
TTCCGCTTCGCTGGGCTCTTCTCTTCTTCTTGCCTCCGCATACCACGCGCCACTG
GGTCGTCTTCATTCAG
CCGCCGCACTGTGCGCTTACCTCCTTTGCCATGCTTGATTAGCACCGGTGGGTTGCT
GAAACCCACCATTTGT
AGCGCCACATCTTCTCTTTCTTCTCGCTGTCCACGATTACCTCTGGTGATGGCGGG
CGCTCGGGCTTGGGAG
AAGGGCGCTTCTTTTTCTTCTTGGGCGCAATGGCCAAATCCGCCGCCGAGGTCGAT
GGCCGCGGGCTGGGTGT
GCGCGCACCAAGCGCGTCTTGATGAGTCTTCTCGTCCCTCGGACTCGATACGCC
GCCTCATCCGCTTTTT
GGGGGCGCCCGGGGAGGCGGCGGCGACGGGGACGGGGACGACACGTCCTCCATG
GTTGGGGGACGTCGCGCCG
CACCGGTCCGCGCTCGGGGGTGGTTTCGCGCTGCTCCTCTTCCCGACTGGCCATTT
CCTTCTCCTATAGGCA
GAAAAAGATCATGGAGTCAGTCGAGAAGAAGGACAGCCTAACCGCCCCCTCTGAG
TTCGCCACCACCGCCTCC
ACCGATGCCGCAACGCGCCTACCACCTTCCCCGTCGAGGCACCCCCGCTTGAGGA
GGAGGAAGTGATTATCG
AGCAGGACCCAGGTTTTGTAAGCGAAGACGACGAGGACCGCTCAGTACCAACAGA
GGATAAAAAGCAAGACCA
GGACAACGCAGAGGCAAACGAGGAACAAGTCGGGCGGGGGGACGAAAGGCATGG
CGACTACCTAGATGTGGGA
GACGACGTGCTGTTGAAGCATCTGCAGCGCCAGTGCGCCATTATCTGCGACGCGTT
GCAAGAGCGCAGCGATG

TGCCCTCGCCATAGCGGATGTCAGCCTTGCCTACGAACGCCACCTATTCTCACCG
CGGTACCCCCCAAACG
CCAAGAAAACGGCACATGCGAGCCCAACCCGCGCCTCAACTTCTACCCCGTATTTG
CCGTGCCAGAGGTGCTT
GCCACCTATCACATCTTTTTCCAAAACCTGCAAGATACCCCTATCCTGCCGTGCCAA
CCGCAGCCGAGCGGACA
AGCAGCTGGCCTTGCGGCAGGGCGCTGTCATACCTGATATCGCCTCGCTCAACGAA
GTGCCAAAAATCTTTGA
GGGTCTTGGACGCGACGAGAAGCGCGCGGCAAACGCTCTGCAACAGGAAAACAG
CGAAAATGAAAGTCACTCT
GGAGTGTGGTGGAACTCGAGTTACCGTCGACCTCTAGCTAGAGCTTGGCGTAATC
ATGGTCATAGCTGTTTC
CTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATA
AAGTGTAAGCCTGGGG
TGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTCC
AGTCGGGAAACCTGTCTG
TGCCAGCTGCATTAATGAATCGGCCAACCGCGGGGAGAGGCGGTTTGCCTATTG
GGCGCTCTCCGCTTCCT
CGCTCACTGACTCGCTGCGCTCGGTCGTTCCGGCTGCGGCGAGCGGTATCAGCTCAC
TCAAAGGCGGTAATACG
GTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAGGCCA
GCAAAGGCCAGGAACCGTA
AAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCAC
AAAAATCGACGCTCAAGT
CAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAA
GCTCCCTCGTGCCTCTCC
TGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGT
GGCGCTTTCTCATAGCT
CACGCTGTAGGTATCTCAGTTCGGTGTAGGTGTTTCGCTCCAAGCTGGGCTGTGTG
CACGAACCCCCGTTTCAG

CCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACA
CGACTTATCGCCACTGGC
AGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAG
TTCTTGAAGTGGTGGCCTA
ACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTT
ACCTTCGGAAAAAGAGTT
GGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTGTTGCAA
GCAGCAGATTACGCGCAG
AAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGT
GGAACGAAAACCTCACGTT
AAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAAT
TAAAAATGAAGTTTTAAA
TCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAG
TGAGGCACCTATCTCAGC
GATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTAC
GATACGGGAGGGCTTAC
CATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGAT
TTATCAGCAATAAACCAG
CCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCC
AGTCTATTAATTGTTGCCG
GGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTG
CTACAGGCATCGTGGTGT
CACGCTCGTCGTTTTGGTATGGCTTCATTCAGCTCCGGTTCCTAACGATCAAGGCCA
GTTACATGATCCCCCATG
TTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCCTCCGATCGTTGTCAGAAGTAAGTT
GGCCGCAGTGTTATCACT
CATGGTTATGGCAGCACTGCATAATTCTCTTACTGTTCATGCCATCCGTAAGATGCTT
TTCTGTGACTGGTGAGT
ACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCCG
GCGTCAATACGGGATAAT

ACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGG
GCGAAAACCTCTCAAGGAT
CTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTT
CAGCATCTTTTACTTTCA
CCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAA
TAAGGGCGACACGGAAATGT
TGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT
CTCATGAGCGGATACAT
ATTTGAATGTATTTAGAAAAATAACAAATAGGGGTTCGCGCACATTTCCCGAA
AAGTGCCACCTGACGTC
SEQ ID NO:48 是 CP 1606 REV 序列 (3855 个碱基对)
AGCGCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAG
CTGGCACGACAGGTTTCCCG
ACTGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTA
GGCACCCAGGCTTTACTACT
TTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTACACA
GGAAACAGCTATGACCATG
ATTACCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGC
TGGAATTCGCCCTTGACGCG
GCCTGTCCGGCCAATCAGTTTTCCGGCAAGTCTTGCGGCATGTTCTTCTCTGAAGG
CGCAAAGGCTCAGGTGGCT
TTTAAGCAGATCAAGGCTTTTATGCAGGCGCTGTATCCTAACGCCAGACCGGGC
ACGGTCACCTTTTGATGCCA
CTACGGTGCAGTGCAACTCAAAGCCTGGGCACGCGCCCTTTTGGGAAGGCAGC
TACCAAAGTTGACTCCGTTT
GCCCTGAGCAACGCGGAGGACCTGGACGCGGATCTGATCTCCGACAAGAGCGTG
CTGGCCAGCGTGCACCACCCG
GCGCTGATAGTGTCCAGTGCTGCAACCCTGTGTATCGCAACTCGCGCGCGCAGG
GCGGAGGCCCAACTGCGAC
TTCAAGATATCGGCGCCCCGACCTGCTAAACGCGTTGGTGATGGTGCGCAGCCTGT

GGAGTGAAAACCTTCACCGAG
CTGCCGCGGATGGTTGTGCCTGAGTTTAAGTGGAGCACTAAACACCAGTATCGCA
ACGTGTCCCTGCCAGTGGCG
CATAGCGATGCGCGGCAGAACCCCTTTGATTTTTAAACGGCGCAGACGGCAAGGG
TGGGGGTAAATAATCACCCG
AGAGTGTACAAATAAAAGCATTTCCTTTATTGAAAGTGTCTCTAGTAGCTAGCG
GGAGGGAGGTCCTGGTCTAG
ACATTCAGAGCTTGTAGAATTTCTGCATCCAGGTCAGAAGAATGTCCACTTCCCCA
AGGGCTTTGGTCAGAGCTG
CTTCTACGTCCAAGTGTGAAATGCTCTCCGGAATAGCAGAAACCGCCTGTGTGCA
CTGTCTCTGATGGAAAACA
TCTCATTTTCTTACTGGGTTGCAGTTGTGACACGATGAGAACAAGTTGTTGGCC
AGAGTAGAGAATGACTTCA
GAGTCTGACTTCAACTGTTCTATTGTGGTGGTTTTTAAAAACAGTTTCAAGTAG
AACTCCAGCAGGGTGTGGA
CAAGGTAACAGCTCTCAGCATCCGAGACGTTCTGCAGAACCTCCTGCTGCAGCAG
CCGGGCACTCGTGATGTTAT
CCTGAGCTTGCATAGTGTCTTTCACAGCCCAGAAGGCTTCCCACAGTTTCTGGGGA
ACAACCCCTTCACTTGGC
AGGGCCCAAAGTGAATTCTTGGCCCTGGGCCCCTGATACCTGGCTCCAGAGAAG
CAGGGTAAAACCCAGGCAAG
GGAGCACAACCATCTGCATTTGAGAGGCTGTCGCCAGCAAAGGAGGGCAGAAGG
GTCTGGCTAAAGTCCACAGGC
TTTGCAGCCTCTGTTGAAAATTCATGATGGCCCTCCTACCGCCTGGATGAATAACA
ATAGTACGAGTCATAAGTA
ATTATTTTACATGTTTTTCAAGTGACAAAAGAAGTGGCGCTCCTAATCTGCGCA
CTGTGGCTGCGGAAGTAGG
GCGAGTGGCGCTCCAGGAAGCTGTAGAGCTGTTCCCTGGTTGCGACGCAGGGTGGG
CTGTACCTGGGGACTGTTGA
GCATGGAGTTGGGTACCCCGGTAAAAAGGGCGAATTCTGCAGATATCCATCACAC

TGGCGGCCGCTCGAGCATGC
ATCTAGAGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGA
GCTTGGCGGCCGAATGGGCTG
ACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTTCGCAGCGCATCGCCTTC
TATCGCCTTCTTGACGAGT
TCTTCTGAATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTGCGCCCTTA
TTCCCTTTTTTGCGGCATT

TTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAA
GATCAGTTGGGTGCACGAGT
GGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCC
GAAGAACGTTTTCCAATGAT
GAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGC
AAGAGCAACTCGGTCGCCG
CATACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCAT
CTTACGGATGGCATGACAGT
AAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAAACACTGCGGCCAACTTA
CTTCTGACAACGATCGGAGG
ACCGAAGGAGCTAACCGCTTTTTTGCAACAACATGGGGGATCATGTAACCTCGCCTT
GATCGTTGGGAACCGGAGCT
GAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGC
AACAACGTTGCGCAAACCTATT
AACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAG
GCGGATAAAGTTGCAGGACC
ACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCG
GTGAGCGTGGGTCTCGCGG
TATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACA
CGACGGGGAGTCAGGCAAC
TATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCAT
TGGTAACTGTCAGACCAAGT

TTACTCATATATACTTTAGATTGATTTAAAACCTTCATTTTTTAATTTAAAAGGATCTA
GGTGAAGATCCTTTTTGA
TAATCTCATGCATGACCAAAAATCCCTTAACGTGAGTTTTTCGTTCCACTGAGCGTCA
GACCCCGTAGAAAAGATCA
AAGGATCTTCTTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAA
AAACCACCGCTACCAGCGG
TGGTTTGTGGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTC
AGCAGAGCGCAGATACCAA
ATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCA
CCGCCTACATACCTCGCTC
TGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGG
TTGGACTCAAGACGATAGT
TACCGGATAAGGCGCAGCGGTGCGGGCTGAACGGGGGGTTCGTGCACACAGCCCA
GCTTGGAGCGAACGACCTACA
CCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGG
GAGAAAGGCGGACAGGTATC
CGGTAAGCGGCAGGGTTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAA
ACGCCTGGTATCTTTATAGTC
CTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGG
GGGCGGAGCCTATGGAAAA
ACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCAC
ATGTTCTTTCCTGCGTTAT
CCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGC
CGCAGCCGAACGACCGAGC
GCAGCGAGTCAGTGAGCGAGGAAGCGGAAG
SEQ ID NO:49 is the sequence for OV 1165
CAGGGTAATCATCATCAATAATATACCTTATTTTGGATTGAA
GCCAATATGATAATGAGG
GGGTGGAGTTTGTGACGTGGCGCGGGCGTGGGAACGGGGC
GGGTGACGTAGTAGTGTGG

CGGAAGTGTGATGTTGCAAGTGTGGCGGAACACATGTAAGC
GACGGATGTGGCAAAAGTG
ACGTTTTTGGTGTGCGCCGGTGTACACAGGAAGTGACAATTT
TCGCGCGGTTTTAGGCGG
ATGTTGTAGTAAATTTGGGCGTAACCGAGTAAGATTTGGCCA
TTTTCGCGGGAAAAGTGA
ATAAGAGGAAGTGAAATCTGAATAATTTTGTGTTACTCATAG
CGCGTAATATTTGTCTAG
GGCCGGGATCTCTGCAGGAATTTGATATCAAGCTTATCGATA
CCGTCGAAACTTGTATTAT
TGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAA
TTTCACAAATAAAGCATT
TTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAAT
GTATCTTATCATGTCTG
GATCGATCCGCTAGCGGCGCGCCGTTTCATCCGGACAAAGCC
TGCGCGCGCCCCGCCCCG
CCATTGGCCGTACCGCCCCGCGCCGCGCCCCATCTCGCCCC
TCGCCGCGGGTCCGGCG
CGTTAAAGCCAATAGGAACCGCCGCGTTGTTCCCGTCACGG
CCGGGGCAGCCAATTGTG
GCGGCGCTCGGCGGCTCGTGGCTCTTTCGCGGCAAAAAGGAT
TTGGCGCGTAAAAGTGGC
CGGGACTTTGCAGGCAGCGGCGGCCGGGGGCGGAGCGGGAT
CGAGCCCTCGATGATATCA
GATCAAACGATATCACCGGTCGACTGAAAATGAGACATATTA
TCTGCCACGGAGGTGTTA
TTACCGAAGAAATGGCCGCCAGTCTTTTGGACCAGCTGATCG
AAGAGGTAAGTGGCTGATA
ATCTTCCACCTCCTAGCCATTTTGAACCACCTACCCTTCACGA
ACTGTATGATTTAGACG

TGACGGCCCCGAAGATCCCAACGAGGAGGCGGTTTCGCAG
ATTTTTCCCGACTCTGTAA
TGTTGGCGGTGCAGGAAGGGATTGACTTACTCACTTTTCCGC
CGGCGCCCGGTTCTCCGG
AGCCGCCTCACCTTTCCCGGCAGCCCGAGCAGCCGGAGCAGA
GAGCCTTGGGTCCGGTTT=
CTATGCCAAACCTTGTACCGGAGGTGATCGATCTTACCTGCC
ACGAGGCTGGCTTTCCAC
CCAGTGACGACGAGGATGAAGAGGGTGAGGAGTTTGTGTTA
GATTATGTGGAGCACCCCG
GGCACGGTTGCAGGTCTTGTCAATTATCACCGGAGGAATACGG
GGGACCCAGATATTATGT
GTTTCGCTTTGCTATATGAGGACCTGTGGCATGTTTGTCTACAG
TAAGTGAAAATTATGGG
CAGTGGGTGATAGAGTGGTGGGTTTGGTGTGGTAATTTTTTTT
TTAATTTTTACAGTTT
GTGGTTTAAAGAATTTTGTATTGTGATTTTTTTAAAAGGTCCT
GTGTCTGAACCTGAGCC
TGAGCCCGAGCCAGAACCGGAGCCTGCAAGACCTACCCGCC
GTCCTAAAATGGCGCCTGC
TATCCTGAGACGCCCACATCACCTGTGTCTAGAGAATGCAA
TAGTAGTACGGATAGCTG
TGACTCCGGTCCTTCTAACACACCTCCTGAGATACACCCGGT
GGTCCCCTGTGCCCAT
TAAACCAGTTGCCGTGAGAGTTGGTGGGCGTCGCCAGGCTGT
GGAATGTATCGAGGACTT
GCTTAACGAGCCTGGGCAACCTTTGGACTTGAGCTGTAAACG
CCCCAGGCCATAAGGTGT
AAACCTGTGATTGCGTGTGTGGTTAACGCCTTTGTTTGCTGAA
TGAGTTGATGTAAGTTT

AATAAAGGGTGAGATAATGTTTAACTTGCATGGCGTGTTAAA
TGGGGCGGGGCTTAAA GG
GTATATAATGCGCCGTGGGCTAATCTTGGTTACATCTGACCTC
ATGGAGGCTTGGGAGTG
TTTGAAGATTTTTCTGCTGTGCGTAACTTGCTGGAACAGAG
CTCTAACAGTACCTCTTG
GTTTTGGAGGTTTCTGTGGGGCTCATCCCAGGCAAAGTT'AGT
CTGCAGA A TTAAGGAGGA
TTACAAGTGGGAATTTGAAGAGCTTTTGAAATCCTGTGGTGA
GCTGTTTGATTCTTTGAA
TCTGGGTCACCAGGCGCTTTTCCAAGAGAAGGTCATCAAGAC
TTTGATTTTTCCACACC
GGGGCGCGCTGCGGCTGCTGTTGCTTTTTTGAGTTTTATAAAG
GATA A ATGGAGCGAAGA
AACCCATCTGAGCGGGGGGTACCTGCTGGATTTTCTGGCCAT
GCATCTGTGGAGAGCGGT
TGTGAGACACAAGAATCGCCTGCTACTGTTGTCTTCCGTCCG
CCCGGCGATAATACCGAC
GGAGGAGCAGCAGCAGCAGCAGGAGGAAGCCAGGCGGGCGG
CGGCAGGAGCAGAGCCCATG
GAACCCGAGAGCCGGCCTGGACCCTCGGGAATGAATGTTGTA
CAGGTGGCTGAACTGTAT
CCAGAACTGAGACGCATTTTGACAATTACAGAGGATGGGCA
GGGGCTA A AGGGGGTAAAG
AGGGAGCGGGGGGCTTGTGAGGCTACAGAGGAGGCTAGGAA
TCTAGCTTTTAGCTTAATG
ACCAGACACCGTCCTGAGTGTATTACTTTTCAACAGATCAAG
GATAATTGCGCTAATGAG
CTTGATCTGCTGGCGCAGAAGTATTCCATAGAGCAGCTGACC
ACTTACTGGCTGCAGCCA

GGGGATGATTTTGGAGGAGGCTATTAGGGTATATGCAAAGGTG
GCACTTAGGCCAGATTGC
AAGTACAAGATCAGCAAACCTTGTAATATCAGGAATTGTTGC
TACATTTCTGGGAACGGG
GCCGAGGTGGAGATAGATACGGAGGATAGGGTGGCCTTTAG
ATGTAGCATGATAAATATG
TGGCCGGGGGTGCTTGGCATGGACGGGGTGGTTATTATGAAT
GTAAGGTTTACTGGCCCC
AATTTTAGCGGTACGGTTTTCTGGCCAATACCAACCTTATCC
TACACGGTGTAAGCTC
TATGGGTTTAAACAATACCTGTGTGGAAGCCTGGACCGATGTA
AGGGTTCGGGGCTGTGCC
TTTTACTGCTGCTGGAAGGGGGTGGTGTGTCGCCCCAAAAGC
AGGGCTTCAATTAAGAAA
TGCCTCTTTGAAAGGTGTACCTTGGGTATCCTGTCTGAGGGTA
ACTCCAGGGTGCGCCAC
AATGTGGCCTCCGACTGTGGTTGCTTCATGCTAGTGAAAAGC
GTGGCTGTGATTAAGCAT
AACATGGTATGTGGCAACTGCGAGGACAGGGCCTCTCAGATG
CTGACCTGCTCGGACGGC
AACTGTCACCTGCTGAAGACCATTCACGTAGCCAGCCACTCT
CGCAAGGCCTGGCCAGTG
TTTGAGCATAACATACTGACCCGCTGTTCTTGCATTTGGGTA
ACAGGAGGGGGGTGTTT
CTACCTTACCAATGCAATTTGAGTCACACTAAGATATTGCTTG
AGCCCGAGAGCATGTCC
AAGGTGAACCTGAACGGGGTGTGTTGACATGACCATGAAGATC
TGGAAGGTGCTGAGGTAC
GATGAGACCCGCACCAGGTGCAGACCCTGCGAGTGTGGCGG
TAAACATATTAGGAACCAG

CCTGTGATGCTGGATGTGACCGAGGAGCTGAGGCCCGATCAC
TTGGTGCTGGCCTGCACC
CGCGCTGAGTTTGGCTCTAGCGATGAAGATACAGATTGAGGT
ACTGAAATGTGTGGGCGT
GGCTTAAGGGTGGGAAAGAATATATAAGGTGGGGGTCTTAT
GTAGTTTTGTATCTGTTTT
GCAGCAGCCGCCGCCGCCATGAGCACCAACTCGTTTGATGGA
AGCATTGTGAGCTCATAT
TTGACAACGCGCATGCCCCCATGGGCCGGGGTGCGTCAGAAT
GTGATGGGCTCCAGCATT
GATGGTCGCCCCGTCCTGCCCGCAAACCTCTACTACCTTGACCT
ACGAGACCGTGTCTGGA
ACGCCGTTGGAGACTGCAGCCTCCGCCGCCGCTTCAGCCGCT
GCAGCCACCGCCCGCGGG
ATTGTGACTGACTTTGCTTTCCTGAGCCCGCTTGCAAGCAGTG
CAGCTTCCCGTTCATCC
GCCCGCGATGACAAGTTGACGGCTCTTTTGGCACAATTGGAT
TCTTTGACCCGGGAACTT
AATGTCGTTTCTCAGCAGCTGTTGGATCTGCGCCAGCAGGTTT
CTGCCCTGAAGGCTTCC
TCCCCTCCCAATGCGGTTTAAACATAAATAAAAAACCAGAC
TCTGTTTGGATTTGGATC
AAGCAAGTGTCTTGCTGTCTTTATTTAGGGGTTTTGCGCGCGC
GGTAGGCCCGGGACCAG
CGGTCTCGGTGTTGAGGGTCTGTGTATTTTTTCCAGGACGT
GGTAAAGGTGACTCTGG
ATGTTTCAGATACATGGGCATAAGCCCGTCTCTGGGGTGGAGG
TAGCACCCTGCAGAGCT
TCATGCTGCGGGGTGGTGTGTAGATGATCCAGTCGTAGCAG
GAGCGCTGGGCGTGGTGC

CTAAAAATGTCTTTCAGTAGCAAGCTGATTGCCAGGGGCAGG
CCCTTGGTGTAAGTGTTT
ACAA AGCGGTTAAGCTGGGATGGGTGCATACGTGGGGATAT
GAGATGCATCTTGGACTGT
ATTTTTAGGTTGGCTATGTTCCAGCCATATCCCTCCGGGGAT
TCATGTTGTGCAGAACC
ACCAGCACAGTGTATCCGGTGCACCTGGGAAATTTGTCATGT
AGCTTAGAAGGA AA TGCG
TGGAAGA ACTTGGAGACGCCCTTGTGACCTCCAAGATTTTCC
ATGCATTCGTCCATAATG
ATGGCAATGGGCCACGGGCGGCGCCTGGGCGAAGATATT
TCTGGGATCACTAACGTCA
TAGTTGTGTTCCAGGATGAGATCGTCATAGGCCATTTTTACA
AAGCGCGGGCGGAGGGTG
CCAGACTGCGGTATAATGGTTCCATCCGGCCCAGGGGCGTAG
TTACCCTCACAGATTTGC
ATTTCCCACGCTTTGAGTTCAGATGGGGGGATCATGTCTACCT
GCGGGGCGATGAAGAAA
ACGGTTTCCGGGGTAGGGGAGATCAGCTGGGAAGAAAGCAG
GTTCTGAGCAGCTGCGAC
TTACCGCAGCCGGTGGGCCCATAAATCACACCTATTACCGGG
TGCAACTGGTAGTTAAGA
GAGCTGCAGCTGCCGTCATCCCTGAGCAGGGGGGCGCACTTCG
TTAAGCATGTCCCTGACT
CGCATGTTTTCCCTGACCAAATCCGCCAGAAGGCGCTCGCCG
CCCAGCGATAGCAGTTCT
TGCAAGGAAGCAAAGTTTTTCAACGGTTTGAGACCGTCCGCC
GTAGGCATGCTTTTGAGC
GTTTGACCAAGCAGTTCAGGCGGTCCCACAGCTCGGTCACC
TGCTCTACGGCATCTCGA

TCCAGCATATCTCCTCGTTTCGCGGGTTGGGGCGGCTTTCGCT
GTACGGCAGTAGTCGGT
GCTCGTCCAGACGGGCCAGGGTCATGTCTTCCACGGGCGCA
GGGTCCTCGTCAGCGTAG
TCTGGGTCACGGTGAAGGGGTGCGCTCCGGGCTGCGCGCTGG
CCAGGGTGCGCTTGAGGC
TGGTCCTGCTGGTGCTGAAGCGCTGCCGGTCTTCGCCCTGCG
CGTCGGCCAGGTAGCATT
TGACCATGGTGTACATAGTCCAGCCCCTCCGCGGCGTGGCCCT
TGGCGCGCAGCTTGCCCT
TGGAGGAGGCGCCGCACGAGGGGCAGTGCAGACTTTTGAGG
GCGTAGAGCTTGGGCGCGA
GAAATACCGATTCCGGGGAGTAGGCATCCGCGCCGCAGGCC
CCGCAGACGGTCTCGCATT
CCACGAGCCAGGTGAGCTCTGGCCGTTCCGGGTCAAAAACCA
GGTTTCCCCCATGCTTTT
TGATGCGTTTCTTACCTCTGGTTTCCATGAGCCGGTGTCCACG
CTCGGTGACGAAAAGGC
TGTCGGTGTCCCCGTATACAGACTTGAGAGGCCTGTCCTCGA
GCGGTGTTCCGCGGTCCT
CCTCGTATAGAACTCGGACCACTCTGAGACAAAGGCTCGCG
TCCAGGCCAGCACGAAGG
AGGCTAAGTGGGAGGGGTAGCGGTCGTTGTCCACTAGGGGG
TCCACTCGCTCCAGGGTGT
GAAGACACATGTCGCCCTCTTCGGCATCAAGGAAGGTGATTG
GTTTGTAGGTGTAGGCCA
CGTGACCGGGTGTTCCTGAAGGGGGGCTATAAAAGGGGGTG
GGGGCGCGTTCGTCCTCAC
TCTCTCCGCATCGCTGTCTGCGAGGGCCAGCTGTTGGGGTG
AGTACTCCCTCTGAAAAG

CGGGCATGACTTCTGCGCTAAGATTGTCAGTTTCCAAAAACG
AGGAGGATTTGATATTCA
CCTGGCCC GCGGTGATGCCTTTGAGGGTGGCCGCATCCATCT
GGTCAGAAAAGACAATCT
TTTTGTTGTCAAGCTTGGTGGCAAACGACCCGTAGAGGGCGT
TGGACAGCAACTTGGCGA
GGAGCGCAGGGTTTGGTTTTTGTGCGGATCGGCGCGCTCCTT
GGCCGCGATGTTTAGCT
GCACGTATTCGCGCGCAACGCACCGCCATTCGGGAAAGACG
GTGGTGCCTCGTCGGGCA
CCAGGTGCACGCGCAACCGCGGTTGTGCAGGGTGACAAGG
TCAACGCTGGTGGCTACCT
CTCCGCGTAGGCGCTCGTTGGTCCAGCAGAGGCGGCCGCCCT
TGCGCGAGCAGAATGGCG
GTAGGGGGTCTAGCTGCGTCTCGTCCGGGGGGTCTGCGTCCA
CGGTAAAGACCCCGGGCA
GCAGGCGCGCTCGAAGTAGTCTATCTTGCATCCTTGCAAGT
CTAGCGCCTGCTGCCATG
CGCGGGCGGCAAGCGCGCTCGTATGGGTTGAGTGGGGGA
CCCATGGCATGGGGTGGG
TGAGCGCGGAGGCGTACATGCCGCAAATGTCGTAAACGTAG
AGGGGCTCTCTGAGTATTC
CAAGATATGTAGGGTAGCATCTTCCACCGCGGATGCTGGCGC
GCACGTAATCGTATAGTT
CGTGCGAGGGAGCGAGGAGGTCGGGACCGAGGTTGCTACGG
GCGGGCTGCTCTGCTCGGA
AGACTATCTGCCTGAAGATGGCATGTGAGTTGGATGATATGG
TTGGACGCTGGAAGACGT
TGAAGCTGGCGTCTGTGAGACCTACCGCGTCACGCACGAAGG
AGGCGTAGGAGTCGCGCA

GCTTGTTGACCAGCTCGGCGGTGACCTGCACGTCTAGGGCGC
AGTAGTCCAGGGTTTCCT
TGATGATGTCATACTTATCCTGTCCCTTTTTTTTCCACAGCTC
GCGGTTGAGGACAAACT
CTTCGCGGTCTTCCAGTACTCTTGGATCGGAAACCCGTCGGC
CTCCGAACGGTAAGAGC
CTAGCATGTAGAACTGGTTGACGGCCTGGTAGGCGCAGCATC
CCTTTTCTACGGGTAGCG
CGTATGCCTGCGCGGCCTTCCGGAGCGAGGTGTGGGTGAGCG
CAAAGGTGTCCCTGACCA
TGACTTTGAGGTACTGGTATTTGAAGTCAGTGTCTCGTCGCATCC
GCCCTGCTCCCAGAGCA
AAAAGTCCGTGCGCTTTTTGGAACGCGGATTTGGCAGGGCGA
AGGTGACATCGTTGAAGA
GTATCTTTCCCGCGCGAGGCATAAAGTTGCGTGTGATGCGGA
AGGGTCCCGGCACCTCGG
AACGGTTGTTAATTACCTGGGCGGCGAGCACGATCTCGTCAA
AGCCGTTGATGTTGTGGC
CCACAATGTAAAGTTCCAAGAAGCGCGGGATGCCCTTGATGG
AAGGCAATTTTTTAAGTT
CCTCGTAGGTGAGCTCTTCAGGGGAGCTGAGCCCGTGCTCTG
AAAGGGCCCAGTCTGCAA
GATGAGGGTTGGAAGCGACGAATGAGCTCCACAGGTCACGG
GCCATTAGCATTTCAGGT
GGTCGCGAAAGGTCCCTAAACTGGCGACCTATGGCCATTTTTT
CTGGGGTGATGCAGTAGA
AGGTAAGCGGGTCTTGTTCAGCGGTCCCATCCAAGGTTTCG
CGGCTAGGTCTCGCGCGG
CAGTCACTAGAGGCTCATCTCCGCCGAACCTTCATGACCAGCA
TGAAGGGCACGAGCTGCT

TCCCAAAGGCCCCCATCCAAGTATAGGTCTCTACATCGTAGG
TGACAAAGAGACGCTCGG
TGCGAGGATGCGAGCCGATCGGGAAGAAGTGGATCTCCCGC
CACCAATTGGAGGAGTGGC
TATTGATGTGGTGAAAGTAGAAGTCCCTGCGACGGGCCGAAC
ACTCGTGCTGGCTTTTGT
AAAAACGTGCGCAGTACTGGCAGCGGTGCACGGGCTGTACA
TCCTGCACGAGGTTGACCT
GACGACCGCGCACAAGGAAGCAGAGTGGGAATTTGAGCCCC
TCGCCTGGCGGGTTTGGCT
GGTGGTCTTCTACTTCGGCTGCTTGTCTTGACCGTCTGGCTG
CTCGAGGGGAGTTACGG
TGGATCGGACCACCACGCCGCGGAGCCCAAAGTCCAGATGT
CCGCGCGCGGCGGTTCGGA
GCTTGATGACAACATCGCGCAGATGGGAGCTGTCCATGGTCT
GGAGCTCCCGCGGCGTCA
GGTCAGGCGGGAGCTCCTGCAGGTTTACCTCGCATAGACGGG
TCAGGGCGCGGGCTAGAT
CCAGGTGATACCTAATTTCCAGGGGCTGGTTGGTGGCGGCGT
CGATGGCTTGCAAGAGGC
CGCATCCCCGCGGCGGCGACTACGGTACCGCGCGGCGGGCGGT
GGGCCGCGGGGGTGTCTT
TGGATGATGCATCTAAAAGCGGTGACGCGGGCGAGCCCCCG
GAGGTAGGGGGGGCTCCGG
ACCCGCCGGGAGAGGGGGCAGGGGCACGTCCGGCGCCGCGCG
CGGGCAGGAGCTGGTGCTG
CGCGCGTAGGTTGCTGGCGAACGCGACGACGCGGCGGTTGAT
CTCCTGAATCTGGCGCCT
CTGCGTGAAGACGACGGGCCCGGTGAGCTTGAGCCTGAAAG
AGAGTTCGACAGAATCAAT

TTCGGTGTGCGTTGACGGCGGCCTGGCGCAAATCTCCTGCAC
GTCTCCTGAGTTGTCTTG
ATAGGCGATCTCGGCCATGAACTGCTCGATCTCTTCTCCTGGAGATCTCCGCGTC
CGGC
TCGCTCCACGGTGGCGGCGAGGTCGTTGGAAATGCGGGCCATGAGCTGCGAGAA
GGCGTT
GAGGCCTCCCTCGTTCCAGACGCGGCTGTAGACCACGCCCCCTTCGGCATCGCGG
GCGCG
CATGACCACCTGCGCGAGATTGAGCTCCACGTGCCGGGCGAAGACGGCGTAGTTT
CGCAG
GCGCTGAAAGAGGTAGTTGAGGGTGGTGGCGGTGTGTTCTGCCACGAAGAAGTAC
ATAAC
CCAGCGTCGCAACGTGGATTGATATCCCCAAGGCCTCAAGGCGCTCCATG
GCCTC
GTAGAAGTCCACGGCGAAGTTGAAAACTGGGAGTTGCGCGCCGACACGGTTAA
CTCCTC
CTCCAGAAGACGGATGAGCTCGGCGACAGTGTGCGGCACCTCGCGCTCAAAGGCT
ACAGG
GGCCTCTTCTTCTTCTTCAATCTCCTCTTCCATAAGGGCCTCCCCTTCTTCTTCTTCT
GG
CGGCGGTGGGGGAGGGGGGACACGGCGGCGACGACGGCGCACCGGGAGGCGGT
CGACAAA
GCGCTCGATCATCTCCCCGCGGCGACGGCGCATGGTCTCGGTGACGGCGCGGCCG
TTCTC
GCGGGGGCGCAGTTGGAAGACCCGCCCCGTCATGTCCCGTTATGGGTTGGCGGG
GGGCT
GCCATGCGGCAGGGATACGGCGCTAACGATGCATCTCAACAATTGTTGTGTAGGT
ACTCC
GCCGCCGAGGGACCTGAGCGAGTCCGCATCGACCGGATCGGAAAACCTCTCGAG
AAAGGC

GTCTAACCAGTCACAGTCGCAAGGTAGGCTGAGCACCGTGGCGGGCGGCAGCGG
GCGGCG
GTCGGGGTTGTTTCTGGCGGAGGTGCTGCTGATGATGTAATTAAGTAGGCGGTC
TTGAG
ACGGCGGATGGTCGACAGAAGCACCATGTCCTTGGGTCCGGCCTGCTGAATGCGC
AGGCG
GTCGGCCATGCCCCAGGCTTCGTTTTGACATCGGCGCAGGTCTTTGTAGTAGTCTT
GCAT
GAGCCTTTCTACCGGCACTTCTTCTTCCTTCCTTGTCTGCATCTCTTGCATC
TAT
CGCTGCGGGCGGCGGCGGAGTTTGGCCGTAGGTGGCGCCCTCTTCCTCCCATGCGT
GTGAC
CCCGAAGCCCCTCATCGGCTGAAGCAGGGCTAGGTCCGGCGACAACGCGCTCGGCT
AATAT
GGCCTGCTGCACCTGCGTGAGGGTAGACTGGAAGTCATCCATGTCCACAAAGCGG
TGGTA
TGCGCCCGTGTTGATGGTGTAAGTGCAGTTGGCCATAACGGACCAGTTAACGGTC
TGGTG
ACCCGGCTGCGAGAGCTCGGTGTACCTGAGACGCGAGTAAGCCCTCGAGTCAAAT
ACGTA
GTCGTTGCAAGTCCGCACCAGGTAAGTGGTATCCCACCAAAAAGTGCGGCGGCGGC
TGGCG
GTAGAGGGGCCAGCGTAGGGTGGCCGGGGCTCCGGGGGCGAGATCTTCCAACAT
AAGGCG
ATGATATCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAG
GCGCG
CGGAAAGTCGCGGACGCGGTTCCAGATGTTGCGCAGCGGCAAAAAGTGCTCCATG
GTCGG
GACGCTCTGGCCGGTCAGGCGCGCGCAATCGTTGACGCTCTAGACCGTGCAAAAG
GAGAG

CCTGTAAGCGGGCACTCTTCCGTGGTCTGGTGGATAAATTCGCAAGGGTATCATG
GCGGA
CGACCGGGGTTTCGAGCCCCGTATCCGGCCGTCCGCCGTGATCCATGCGGTTACCG
CCCGC
GTGTGGAACCCAGGTGTGCGACGTCAGACAACGGGGGAGTGCTCCTTTTGGCTTC
CTTCC
AGGCGCGGCGGCTGCTGCGCTAGCTTTTTTGGCCACTGGCCGCGCGCAGCGTAAG
CGGTT
AGGCTGGAAAGCGAAAGCATTAAGTGGCTCGCTCCCTGTAGCCGGAGGGTTATTT
TCCAA
GGGTTGAGTCGCGGGACCCCCGGTTCGAGTCTCGGACCGGCCGACTGCGGGCAA
CGGGG
GTTTGCTCCCCGTCATGCAAGACCCCGCTTGCAAATTCCTCCGAAACAGGGAC
GAGCC
CCTTTTTTGCTTTTCCCAGATGCATCCGGTGCTGCGGCAGATGCGCCCCCTCCTC
AGCA
GCGGCAAGAGCAAGAGCAGCGGCAGACATGCAGGGCACCCCTCCCCTCCTCCTACC
GCGTC
AGGAGGGGCGACATCCGCGGTTGACGCGGCAGCAGATGGTGATTACGAACCCCC
GCGGCG
CCGGGCCCGGCACTACCTGGACTTGGAGGAGGGCGAGGGCCTGGCGCGGCTAGG
AGCGCC
CTCTCCTGAGCGGTACCCAAGGGTGCAGCTGAAGCGTGATACGCGTGAGGCGTAC
GTGCC
GCGGCAGAACCTGTTTCGCGACCGCGAGGGAGAGGCCGAGGAGATGCGGGA
TCGAAA
GTTCCACGCAGGGCGCGAGCTGCGGCATGGCCTGAATCGCGAGCGGTTGCTGCGC
GAGGA
GGACTTTGAGCCCGACGCGGAACCGGGATTAGTCCCGCGCGCGCACACGTGGCG
GCCGC

CGACCTGGTAACCGCATACGAGCAGACGGTGAACCAGGAGATTAAC TTTCAAAA
AAGCTT
TAACAACCACGTGCGTACGCTTGTGGCGCGGAGGAGGTGGCTATAGGACTGATG
CATCT
GTGGGACTTTGTAAGCGCGCTGGAGCAAACCCAAATAGCAAGCCGCTCATGGCG
CAGCT
G TTCCTTATAGTGCAGCACAGCAGGGACAACGAGGCATTCAGGGATGCGCTGCTA
AACAT
AGTAGAGCCCGAGGGCCGCTGGCTGCTCGATTTGATAAACATCCTGCAGAGCATA
GTGGT
GCAGGAGCGCAGCTTGAGCCTGGCTGACAAGGTGGCCGCATCAACTATTCCATG
CTTAG
CCTGGGCAAGTTTTACGCCGCAAGATATACCATACCCCTTACGTTCCCATAGACA
AGGA
GGTAAAGATCGAGGGGTTCTACATGCGCATGGCGCTGAAGGTGCTTACCTTGAGC
GACGA
CCTGGGCGTTTATCGCAACGAGCGCATCCACAAGGCCGTGAGCGTGAGCCGGCGG
CGCGA
GCTCAGCGACCGCGAGCTGATGCACAGCCTGCAAAGGGCCCTGGCTGGCACGGG
CAGCGG
CGATAGAGAGGCCGAGTCCTACTTTGACGCGGGCGCTGACCTGCGCTGGGCCCA
AGCCG
ACGCGCCCTGGAGGCAGCTGGGGCCGACCTGGGCTGGCGGTGGCACCCGCGCG
CGCTGG
CAACGTCGGCGGCGTGGAGGAATATGACGAGGACGATGAGTACGAGCCAGAGGA
CGGCGA
GTACTAAGCGGTGATGTTTCTGATCAGATGATGCAAGACGCAACGGACCCGGCGG
TGCGG
GCGGCGCTGCAGAGCCAGCCGTCCGGCCTTA ACTCCACGGACGACTGGCGCCAGG
TCATG

GACCGCATCATGTCGCTGACTGCGCGCAATCCTGACGCGTTCGGGCAGCAGCCGC
AGGCC
AACCGGCTCTCCGCAATTCTGGAAGCGGTGGTCCCGGCGCGCGCAAACCCACGC
ACGAG
AAGGTGCTGGCGATCGTAAACGCGCTGGCCGAAAACAGGGCCATCCGGCCCGAC
GAGGCC
GGCCTGGTCTACGACGCGCTGCTTCAGCGCGTGGCTCGTTACAACAGCGGCAACG
TGCAG
ACCAACCTGGACCGGCTGGTGGGGATGTGCGCGAGGCCGTGGCGCAGCGTGAG
CGCGCG
CAGCAGCAGGGCAACCTGGGCTCCATGGTTGCACTAAACGCCTTCCTGAGTACAC
AGCCC
GCCAACGTGCCGCGGGGACAGGAGGACTACACCAACTTTGTGAGCGCACTGCGG
CTAATG
GTGACTGAGACACCGCAAAGTGAGGTGTACCAGTCTGGGCCAGACTATTTTTTCC
AGACC
AGTAGACAAGGCCTGCAGACCGTAAACCTGAGCCAGGCTTTCAAAACTTGCAGG
GGCTG
TGGGGGGTGCGGGCTCCACAGGCGACCGCGCGACCGTGTCTAGCTTGCTGACGC
CCAAC
TCGCGCCTGTTGCTGCTGCTAATAGCGCCCTTCACGGACAGTGGCAGCGTGTCCCG
GGAC
ACATACCTAGGTCACTTGCTGACACTGTACCGCGAGGCCATAGGTCAGGCGCATG
TGGAC
GAGCATACTTCCAGGAGATTACAAGTGTGAGCCGCGCGCTGGGGCAGGAGGAC
ACGGGC
AGCCTGGAGGCAACCCTAAACTACCTGCTGACCAACCGGCGGCAGAAGATCCCCT
CGTTG
CACAGTTTAAACAGCGAGGAGGAGCGCATTTTTCGCTACGTGCAGCAGAGCGTGA
GCCTT

AACCTGATGCGCGACGGGGTAACGCCAGCGTGGCGCTGGACATGACCGCGCGC
AACATG
GAACCGGGCATGTATGCCTCAAACCGGCCGTTTATCAACCGCCTAATGGACTACT
TGCAT
CGCGCGGCCGCCGTGAACCCCGAGTATTTACCAATGCCATCTTGAACCCGCACT
GGCTA
CCGCCCCCTGGTTTCTACACCGGGGGATTTCGAGGTGCCCGAGGGTAACGATGGAT
TCCTC
TGGGACGACATAGACGACAGCGTGTTTTCCCGCAACCGCAGACCCTGCTAGAGT
TGCAA
CAGCGCGAGCAGGCAGAGGCGGCGCTGCGAAAGGAAAGCTTCCGCAGGCCAAGC
AGCTTG
TCCGATCTAGGCGCTGCGGCCCGCGGTCAGATGCTAGTAGCCCATTTCCAAGCTT
GATA
GGGTCTCTTACCAGCACTCGCACCACCCGCCCGCCTGCTGGGCGAGGAGGAGT
ACCTA
AACAACCTCGCTGCTGCAGCCGCAGCGCGAAAAAACCTGCCTCCGGCATTTCCTCA
ACAAC
GGGATAGAGAGCCTAGTGGACAAGATGAGTAGATGGAAGACGTACGCGCAGGAG
CACAGG
GACGTGCCAGGCCCGCGCCCGCCACCCGTCGTCAAAGGCACGACCGTCAGCGGG
GTCTG
GTGTGGGAGGACGATGACTCGGCAGACGACAGCAGCGTCCTGGATTTGGGAGGG
AGTGGC
AACCCGTTTGCACCTTCGCCCCAGGCTGGGGAGAATGTTTTAAAAAAAAAAAA
GCATG
ATGCAAATAAAAACTACCAAGGCCATGGCACCGAGCGTTGGTTTTCTTGAT
TCCCC
TTAGTATGCGGCGCGCGGCGATGTATGAGGAAGGTCCTCCTCCCTCCTACGAGAG
TGTGG

TGAGCGCGGCGCCAGTGGCGGGCGGCTGGGTTCTCCCTTCGATGCTCCCCTGGA
CCCGC
CGTTTGTGCCTCCGCGGTACCTGCGGCCTACCGGGGGGAGAAACAGCATCCGTTA
CTCTG
AGTTGGCACCCCTATTCGACACCACCCGTGTGTACCTGGTGGACAACAAGTCAAC
GGATG
TGGCATCCCTGAACTACCAGAACGACCACAGCAACTTTCTGACCACGGTCATTCA
AAACA
ATGACTACAGCCCAGGGGGAGGCAAGCACACAGACCATCAATCTTGACGACCGGT
CGCACT
GGGGCGGCGACCTGAAAACCATCCTGCATACCAACATGCCAAATGTGAACGAGTT
CATGT
TTACCAATAAGTTTAAGGCGCGGGTGATGGTGTGCGGCTTGCCTACTAAGGACAA
TCAGG
TGGAGCTGAAATACGAGTGGGTGGAGTTCACGCTGCCCCGAGGGCAACTACTCCGA
GACCA
TGACCATAGACCTTATGAACAACGCGATCGTGGAGCACTACTTGAAAGTGGGCAG
ACAGA
ACGGGGTTCTGGAAAGCGACATCGGGGTAAAGTTTGACACCCGCAACTTCAGACT
GGGGT
TTGACCCCGTCACTGGTCTTGTCATGCCTGGGGTATATACAAACGAAGCCTTCCAT
CCAG
ACATCATTTTGCTGCCAGGATGCGGGGTGGACTTCACCCACAGCCGCCTGAGCAA
CTTGT
TGGGCATCCGCAAGCGGCAACCCTTCCAGGAGGGCTTTAGGATCACCTACGATGA
TCTGG
AGGGTGGTAACATTCCCGCACTGTTGGATGTGGACGCCTACCAGGCGAGCTTGAA
AGATG
ACACCGAACAGGGCGGGGGTGGCGCAGGCGGCAGCAACAGCAGTGGCAGCGGCG
CGGAAG

AGAACTCCAACGCGGCAGCCGCGGCAATGCAGCCGGTGGAGGACATGAACGATC
ATGCCA
TTCGCGGCGACACCTTTGCCACACGGGCTGAGGAGAAGCGCGCTGAGGCCGAAG
CAGCGG
CCGAAGCTGCCGCCCCCGCTGCGCAACCCGAGGTCGAGAAGCCTCAGAAGAAAC
CGGTGA
TCAAACCCCTGACAGAGGACAGCAAGAAACGCAGTTACAACCTAATAAGCAATG
ACAGCA
CCTTCACCCAGTACCGCAGCTGGTACCTTGCATACAACCTACGGCGACCCTCAGAC
CGGAA
TCCGCTCATGGACCCTGCTTTGCACTCCTGACGTAACCTGCGGCTCGGAGCAGGTC
TACT
GGTCGTTGCCAGACATGATGCAAGACCCCGTGACCTTCCGCTCCACGCGCCAGAT
CAGCA
ACTTTCGGTGGTGGGCGCCGAGCTGTTGCCCGTGCACTCCAAGAGCTTCTACAA
CGACC
AGGCCGTCTACTCCCAACTCATCCGCCAGTTTACCTCTCTGACCCACGTGTTCAAT
CGCT
TTCCCGAGAACCAGATTTTGGCGCGCCCGCCAGCCCCACCATCACCACCGTCAG
TGAAA
ACGTTCTGCTCTCACAGATCACGGGACGCTACCGCTGCGCAACAGCATCGGAGG
AGTCC
AGCGAGTGACCATTACTGACGCCAGACGCCGCACCTGCCCCTACGTTTACAAGGC
CCTGG
GCATAGTCTCGCCGCGCGTCCTATCGAGCCGCACTTTTTGAGCAAGCATGTCCATC
CTTA
TATCGCCCAGCAATAACACAGGCTGGGGCCTGCGCTTCCCAAGCAAGATGTTTGG
CGGGG
CCAAGAAGCGCTCCGACCAACACCCAGTGCGCGTGCGCGGGCACTACCGCGCGCC
CTGGG

GCGCGCACAAACGCGGCCGCACTGGGCGCACACCAGTCGATGACGCCATCGACG
CGGTGG
TGGAGGAGGCGCGCAACTACACGCCACGCCGCCACCAGTGTCCACAGTGGACG
CGGCCA
TTCAGACCGTGGTGCGCGGAGCCCGGCGCTATGCTAAAATGAAGAGACGGCGGA
GGCGCG
TAGCACGTCGCCACCGCCGCCGACCCGGCACTGCCGCCAACGCGCGGGCGGGCGGC
CCTGC
TTAACCGCGCACGTCGCACCGGCCGACGGGCGGCCATGCGGGCCGCTCGAAGGCT
GGCCG
CGGGTATTGTCACTGTGCCCCCAAGTCCAGGCGACGAGCGGCCGCCGACGAGCAGC
CGCGG
CCATTAGTGCTATGACTCAGGGTCGCAGGGGCAACGTGTATTGGGTGCGCGACTC
GGTTA
GCGGCCTGCGCGTGCCCGTGCGCACCCGCCCCCGCGCAACTAGATTGCAAGAAA
AAACT
ACTTAGACTCGTACTGTTGTATGTATCCAGCGGCCGGCGGCCGCAACGAAGCTAT
GTCCA
AGCGCAAAATCAAAGAAGAGATGCTCCAGGTCATCGCGCCGGAGATCTATGGCC
CCCCGA
AGAAGGAAGAGCAGGATTACAAGCCCCGAAAGCTAAAGCGGGTCAAAAAGAAA
AAGAAAG
ATGATGATGATGAACTTGACGACGAGGTGGAAGTCTGTCACGCTACCGCGCCCAG
GCGAC
GGGTACAGTGGAAGGTGCGACGCGTAAAACGTGTTTTGCGACCCGGCACACCAGT
AGTCT
TTACGCCCGGTGAGCGCTCCACCCGCACCTACAAGCGCGTGTATGATGAGGTGTA
CGGCG
ACGAGGACCTGCTTGAGCAGGCCAACGAGCGCCTCGGGGAGTTTGCCTACGGAA
AGCGGC

ATAAGGACATGCTGGCGTTGCCGCTGGACGAGGGCAACCCAACACCTAGCCTAAA
GCCCC
TAACACTGCAGCAGGTGCTGCCCCGCGCTTGACCGTCCGAAGAAAAGCGCGGCCT
AAAGC
GCGAGTCTGGTGACTIONGGCACCCACCGTGCAGCTGATGGTACCCAAGCGCCAGCG
ACTGG
AAGATGTCTTGAAAAAATGACCGTGGAACCTGGGCTGGAGCCCGAGGTCCGCGT
GCGGC
CAATCAAGCAGGTGGCGCCGGGACTGGGCGTGCAGACCGTGGACGTTTCAGATAC
CCACTA
CCAGTAGCACCAGTATTGCCACCGCCACAGAGGGCATGGAGACACAAACGTCCCC
GGTTG
CCTCAGCGGTGGCGGATGCCGCGGTGCAGGCGGTGCGTGCGGCCGCGTCCAAGAC
CTCTA
CGGAGGTGCAAACGGACCCGTGGATGTTTCGCGTTTCAGCCCCCGGCGCCCGCG
CGGTT
CGAGGAAGTACGGCGCCGCCAGCGCGCTACTGCCCGAATATGCCCTACATCCTTC
CATTG
CGCCTACCCCGGCTATCGTGGCTACACCTACCGCCCCAGAAGACGAGCAACTAC
CCGAC
GCCGAACCACCACTGGAACCCGCCGCCGCGTCGCCGTCGCCAGCCCGTGTGGC
CCCGA
TTTCCGTGCGCAGGGTGGCTCGCGAAGGAGGCAGGACCCTGGTGTGCCAACAGC
GCGCT
ACCACCCAGCATCGTTTAAAAGCCGGTCTTTGTGGTTCTTGCAGATATGGCCCTC
ACCT
GCCGCCTCCGTTTCCCGGTGCCGGGATTCCGAGGAAGAATGCACCGTAGGAGGGG
CATGG
CCGGCCACGGCCTGACGGGCGGCATGCGTCGTGCGCACCAACCGGCGGCGGCGCG
CGTCGC

ACCGTCGCATGCGCGGCGGTATCCTGCCCTCCTTATTCCACTGATCGCCGCGGCG
ATTG
GCGCCGTGCCCGGAATTGCATCCGTGGCCTTGCAGGCGCAGAGACACTGATTA
AACAA
GTTGCATGTGGAAAAATCAAAATAAAAAGTCTGGACTCTCACGCTCGCTTGGTCC
TGTA
CTATTTTGTAGAATGGAAGACATCAACTTTGCGTCTCTGGCCCCGCGACACGGCTC
GCGC
CCGTTCATGGGAAACTGGCAAGATATCGGCACCAGCAATATGAGCGGTGGCGCCT
TCAGC
TGGGGCTCGCTGTGGAGCGGCATTAATAATTTTCGGTTCCACCGTTAAGAACTATG
GCAGC
AAGGCCTGGAACAGCAGCACAGGCCAGATGCTGAGGGATAAGTTGAAAGAGCAA
AATTTC
CAACAAAAGGTGGTAGATGGCCTGGCCTCTGGCATTAGCGGGGTGGTGGACCTGG
CCAAC
CAGGCAGTGCAAATAAGATTAACAGTAAGCTTGATCCCCGCCCTCCCGTAGAGG
AGCCT
CCACCGGCCGTGGAGACAGTGTCTCCAGAGGGGCGTGGCGAAAAGCGTCCGCGC
CCCGAC
AGGGAAGAACTCTGGTGACGCAAATAGACGAGCCTCCCTCGTACGAGGAGGCA
CTAAAG
CAAGGCCTGCCACCACCCGTCCCATCGCGCCCATGGCTACCGGAGTGCTGGGCC
AGCAC
ACACCCGTAACGCTGGACCTGCCTCCCCCGCCGACACCAGCAGAAACCTGTGC
TGCCA
GGCCCCGACCGCCGTTGTTGTAACCCGTCCTAGCCGCGCGTCCCTGCGCCGCGCCG
CCAGC
GGTCCGCGATCGTTGCGGCCCGTAGCCAGTGGCAACTGGCAAAGCACACTGAACA
GCATC

GTGGGTCTGGGGGTGCAATCCCTGAAGCGCCGACGATGCTTCTGAATAGCTAACG
TGTCG
TATGTGTGTCATGTATGCGTCCATGTCCGCCAGAGGAGCTGCTGAGCCGCCGC
GCGCC
CGCTTTCCAAGATGGCTACCCCTTCGATGATGCCGCAGTGGTCTTACATGCACATC
TCGG
GCCAGGACGCCTCGGAGTACCTGAGCCCCGGGCTGGTGCAGTTTGCCCGCGCCAC
CGAGA
CGTACTTCAGCCTGAATAACAAGTTTAGAAACCCACGGTGGCGCCTACGCACGA
CGTGA
CCACAGACCGGTCCCAGCGTTTGACGCTGCGGTTTCATCCCTGTGGACCGTGAGGA
TACTG
CGTACTCGTACAAGGCGCGGTTACCCTAGCTGTGGGTGATAACCGTGTGCTGGA
CATGG
CTTCCACGTACTTTGACATCCGCGGCGTGCTGGACAGGGGCCCTACTTTTAAGCCC
TACT
CTGGCACTGCCTACAACGCCCTGGCTCCCAAGGGTGCCCCAAATCCTTGCGAATG
GGATG
AAGCTGCTACTGCTCTTGAAATAAACCTAGAAGAAGAGGACGATGACAACGAAG
ACGAAG
TAGACGAGCAAGCTGAGCAGCAAAAACTCACGTATTTGGGCAGGCGCCTTATTC
TGGTA
TAAATATTACAAAGGAGGGTATTCAAATAGGTGTCGAAGGTCAAACACCTAAATA
TGCCG
ATAAAACATTTCAACCTGAACCTCAAATAGGAGAATCTCAGTGGTACGAAACTGA
AATTA
ATCATGCAGCTGGGAGAGTCCTTAAAAAGACTACCCCAATGAAACCATGTTACGG
TTCAT
ATGCAAAACCCACAAATGAAAATGGAGGGCAAGGCATTCTTGTAAGCAACAAA
ATGGAA

AGCTAGAAAGTCAAGTGGAAATGCAATTTTTCTCAACTACTGAGGCGACCGCAGG
CAATG
GTGATAACTTGACTCCTAAAGTGGTATTGTACAGTGAAGATGTAGATATAGAAAC
CCCAG
ACACTCATATTTCTTACATGCCCACTATTAAGGAAGGTAACCTCACGAGAACTAAT
GGGCC
AACAACTCTATGCCCAACAGGCCTAATTACATTGCTTTTAGGGACAATTTATTGGT
CTAA
TGTATTACAACAGCACGGGTAATATGGGTGTTCTGGCGGGCCAAGCATCGCAGTT
GAATG
CTGTTGTAGATTTGCAAGACAGAAACACAGAGCTTTCATACCAGCTTTTGCTTGAT
TCCA
TTGGTGATAGAACCAGGTACTTTTCTATGTGGAATCAGGCTGTTGACAGCTATGAT
CCAG
ATGTTAGAATTATTGAAAATCATGGAACCTGAAGATGAACTTCCAAATTACTGCTTT
CCAC
TGGGAGGTGTGATTAATACAGAGACTCTTACCAAGGTAAAACCTAAAACAGGTCA
GGAAA
ATGGATGGGAAAAAGATGCTACAGAATTTTCAGATAAAAATGAAATAAGAGTTG
GAAATA
ATTTTGCCATGGAAATCAATCTAAATGCCAACCTGTGGAGAAATTCCTGTACTCC
AACA
TAGCGCTGTATTTGCCCGACAAGCTAAAGTACAGTCCTTCCAACGTAAAAATTTCT
GATA
ACCCAAACACCTACGACTACATGAACAAGCGAGTGGTGGCTCCCGGGTTAGTGGA
CTGCT
ACATTAACCTTGGAGCACGCTGGTCCCTTGACTATATGGACAACGTCAACCCATTT
AACC
ACCACCGCAATGCTGGCCTGCGCTACCGCTCAATGTTGCTGGGCAATGGTCGCTA
TGTGC

CCTTCCACATCCAGGTGCCTCAGAAGTTCTTTGCCATTAAAAACCTCCTTCTCCTG
CCGG
GCTCATAACCTACGAGTGGAACCTCAGGAAGGATGTAAACATGGTTCTGCAGAG
CTCCC
TAGGAAATGACCTAAGGGTTGACGGAGCCAGCATTAAAGTTTGATAGCATTTCCT
TTACG
CCACCTTCTTCCCCATGGCCCACAACACCGCCTCCACGCTTGAGGCCATGCTTAGA
AACG
ACACCAACGACCAGTCCTTTAACGACTATCTCTCCGCCGCAACATGCTCTACCT
ATAC
CCGCCAACGCTACCAACGTGCCCATATCCATCCCCTCCCGCAACTGGGCGGCTTTC
CGCG
GCTGGGCCTTCACGCGCCTTAAGACTAAGGAAACCCCATCACTGGGCTCGGGCTA
CGACC
CTTATTACACCTACTCTGGCTCTATACCCTACCTAGATGGAACCTTTTACCTCAAC
CACA
CCTTTAAGAAGGTGGCCATTACCTTTGACTCTTCTGTCAGCTGGCCTGGCAATGAC
CGCC
TGCTTACCCCAACGAGTTTGAAATTAAGCGCTCAGTTGACGGGGAGGGTTACAA
CGTTG
CCCAGTGTAACATGACCAAAGACTGGTTCCTGGTACAAATGCTAGCTAACTACAA
CATTG
GCTACCAGGGCTTCTATATCCCAGAGAGCTACAAGGACCGCATGTACTCCTTCTTT
AGAA
ACTTCCAGCCCATGAGCCGTCAGGTGGTGGATGATACTAAATACAAGGACTACCA
ACAGG
TGGGCATCCTACACCAACACAACAACCTCTGGATTTGTTGGCTACCTTGCCCCACC
ATGC
GCGAAGGACAGGCCTACCCTGCTAACTTCCCCTATCCGCTTATAGGCAAGACCGC
AGTTG

ACAGCATTACCCAGAAAAAGTTTCTTTGCGATCGCACCCCTTTGGCGCATCCCATTCTCCA
GTAAC TTTATGTCCATGGGCGCACTCACAGACCTGGGCCAAAACCTTCTCTACGCCAACT
CCGCCCACGCGCTAGACATGACTTTTGAGGTGGATCCCATGGACGAGCCCACCCTTCTTT
ATGTTTTGTTTGAAGTCTTTGACGTGGTCCGTGTGCACCGGCCGCACCGCGGCGTCATCG
AAACCGTGTACCTGCGCACGCCCTTCTCGGCCGGCAACGCCACAACATAAAGAAGCAAGC
AACATCAACAACAGCTGCCGCCATGGGCTCCAGTGAGCAGGAACTGAAAGCCATTGTCAA
AGATCTTGGTTGTGGGCCATATTTTTTTGGGCACCTATGACAAGCGCTTTCCAGGCTTTGT
TTCTCCACACAAGCTCGCCTGCGCCATAGTCAATACGGCCGGTCGCGAGACTGGGGCGGT
ACACTGGATGGCCTTTGCCTGGAACCCGCACTCAAAAACATGCTACCTCTTTGAGCCCTT
TGGCTTTTCTGACCAGCGACTCAAGCAGGTTTACCAGTTTGAGTACGAGTCACTCCGTGCG
CCGTAGCGCCATTGCTTCTTCCCCGACCGCTGTATAACGCTGGAAAAGTCCACCCAAAG
CGTACAGGGGCCAACTCGGCCGCCTGTGGACTATTCTGCTGCATGTTTCTCCACGCCTT
TGCCAACTGGCCCCAACTCCCATGGATCACAACCCACCATGAACCTTATTACCGGGGT
ACCCAACTCCATGCTCAACAGTCCCCAGGTACAGCCCACCCTGCGTCGCAACCAGGAACA
GCTCTACAGCTTCTGGAGCGCCACTCGCCCTACTTCCGCAGCCACAGTGCGCAGATTAG

GAGCGCCACTTCTTTTTGTCACTTGAAAAACATGTAAAAATAATGTACTAGAGAC
ACTTT
CAATAAAGGCAAATGCTTTTATTTGTACACTCTCGGGTGATTATTTACCCCCACCC
TTGC
CGTCTGCGCCGTTTAAAAATCAAAGGGGTTCTGCCGCGCATCGCTATGCGCCACT
GGCAG
GGACACGTTGCGATACTGGTGTTTAGTGCTCCACTTAAACTCAGGCACAACCATC
GGCGG
CAGCTCGGTGAAGTTTTCACTCCACAGGCTGCGCACCATCACCAACGCGTTTAGC
AGGTC
GGGCGCCGATATCTTGAAGTCGCAGTTGGGGCCTCCGCCCTGCGCGCGAGTTG
CGATA
CACAGGGTTGCAGCACTGGAACACTATCAGCGCCGGTGGTGCACGCTGGCCAGC
ACGCT
CTTGTCGGAGATCAGATCCGCGTCCAGGTCTCCGCGTTGCTCAGGGCGAACGGA
GTCAA
CTTTGGTAGCTGCCTTCCCAAAAAGGGCGCGTGCCAGGCTTTGAGTTGCACTCGC
ACCG
TAGTGGCATCAAAGGTGACCGTGCCCGGTCTGGGCGTTAGGATACAGCGCCTGC
ATAAA
AGCCTTGATCTGCTTAAAAGCCACCTGAGCCTTTGCGCCTTCAGAGAAGAACATG
CCGCA
AGACTTGCCGAAAACCTGATTGGCCGGACAGGCCGCGTCGTGCACGCAGCACCTT
GCGTC
GGTGTTGGAGATCTGCACCACATTTCCGGCCCCACCGGTTCTTCACGATCTTGGCCT
TGCT
AGACTGCTCCTTCAGCGCGCGCTGCCCGTTTTCGCTCGTCACATCCATTTCAATCA
CGTG
CTCCTTATTTATCATAATGCTTCCGTGTAGACACTTAAGCTCGCCTTCGATCTCAG
CGCA

GCGGTGCAGCCACAACGCGCAGCCCGTGGGCTCGTGATGCTTGTAGGTACCTCT
GCAAA
CGACTGCAGGTACGCCTGCAGGAATCGCCCCATCATCGTCACAAAGGTCTTGTTG
CTGGT
GAAGGTCAGCTGCAACCCGCGGTGCTCCTCGTTCAGCCAGGTCTTGCATACGGCC
GCCAG
AGCTTCCACTTGGTCAGGCAGTAGTTTGAAGTTCGCCTTTAGATCGTTATCCACGT
GGTA
CTTGTCATCAGCGCGCGCAGCCTCCATGCCCTTCTCCCACGCAGACACGATCG
GCAC
ACTCAGCGGGTTCATCACCGTAATTTCACTTTCCGCTTCGCTGGGCTCTCCTCTTC
CTC
TTGCGTCCGCATACCACGCGCCACTGGGTGCTCTTCATTCAGCCGCCGCACTGTGC
GCTT
ACCTCCTTTGCCATGCTTGATTAGCACCGGTGGGTTGCTGAAACCCACCATTTGTA
GCGC
CACATCTTCTTTCTTCTCCTCGCTGTCCACGATTACCTCTGGTGATGGCGGGCGCTC
GGG
CTTGGGAGAAGGGCGCTTCTTTTTCTTCTTGGGCGCAATGGCCAAATCCGCCGCCG
AGGT
CGATGGCCGCGGGCTGGGTGTGCGCGGCACCAGCGCGTCTTGTGATGAGTCTTCC
TCGTC
CTCGGACTCGATACGCCGCTCATCCGCTTTTTTGGGGGCGCCCGGGGAGGCGGC
GGCGA
CGGGGACGGGGACGACACGTCTCCATGGTTGGGGGACGTCGCGCCGCACCGCGT
CCGCG
CTCGGGGGTGGTTTCGCGCTGCTCCTCTTCCCGACTGGCCATTCCTTCTCCTATAG
GCA
GAAAAAGATCATGGAGTCAGTCGAGAAGAAGGACAGCCTAACCGCCCCCTCTGA
GTTTCG

CACCACCGCCTCCACCGATGCCGCAACGCGCCTACCACCTTCCCCGTCGAGGCA
CCCCC
GCTTGAGGAGGAGGAAGTGATTATCGAGCAGGACCCAGGTTTTGTAAGCGAAGA
CGACGA
GGACCGCTCAGTACCAACAGAGGATAAAAAGCAAGACCAGGACAACGCAGAGGC
AAACGA
GGAACAAGTCGGGCGGGGGGACGAAAGGCATGGCGACTACCTAGATGTGGGAGA
CGACGT
GCTGTTGAAGCATCTGCAGCGCCAGTGCGCCATTATCTGCGACGCGTTGCAAGAG
CGCAG
CGATGTGCCCTCGCCATAGCGGATGTCAGCCTTGCCTACGAACGCCACCTATTCT
CACC
GCGCGTACCCCCAAACGCCAAGAAAACGGCACATGCGAGCCCAACCCGCGCCT
CAACTT
CTACCCCGTATTTGCCGTGCCAGAGGTGCTTGCCACCTATCACATCTTTTTCCAAA
ACTG
CAAGATACCCTATCCTGCCGTGCCAACCGCAGCCGAGCGGACAAGCAGCTGGCC
TTGCG
GCAGGGCGCTGTCATACCTGATATCGCCTCGCTCAACGAAGTGCCAAAAATCTTT
GAGGG
TCTTGACGCGACGAGAAGCGCGCGGCAAACGCTCTGCAACAGGAAAACAGCGA
AAATGA
AAGTCACTCTGGAGTGTTGGTGGAACTCGAGGGTGACAACGCGCGCCTAGCCGTA
CTAAA
ACGCAGCATCGAGGTCACCCACTTTGCCTACCCGGCACTTAACCTACCCCCAAG
GTCAT
GAGCACAGTCATGAGTGAGCTGATCGTGCGCCGTGCGCAGCCCCTGGAGAGGGAT
GCAAA
TTTGCAAGAACAACAGAGGAGGGCCTACCCGCAGTTGGCGACGAGCAGCTAGC
GCGCTG

GCTTCAAACGCGCGAGCCTGCCGACTTGGAGGAGCGACGCAAACCTAATGATGGCC
GCAGT
GCTCGTTACCGTGGAGCTTGAGTGCATGCAGCGGTTCTTTGCTGACCCGGAGATG
CAGCG
CAAGCTAGAGGAAACATTGCACTACACCTTTCGACAGGGCTACGTACGCCAGGCC
TGCAA
GATCTCCAACGTGGAGCTCTGCAACCTGGTCTCCTACCTTGGAATTTTGCACGAAA
ACCG
CCTTGGGCAAACGTGCTTCATTCCACGCTCAAGGGCGAGGGCGCGCCGCGACTAC
GTCCG
CGACTGCGTTTACTTATTTCTATGCTACACCTGGCAGACGGCCATGGGCGTTTGGC
AGCA
GTGCTTGGAGGAGTGCAACCTCAAGGAGCTGCAGAACTGCTAAAGCAAACCTT
GAAGGA
CCTATGGACGGCCTTCAACGAGCGCTCCGTGGCCGCGCACCTGGCGGACATCATT
TTCCC
CGAACGCCTGCTTAAAACCCTGCAACAGGGTCTGCCAGACTTCACCAGTCAAAGC
ATGTT
GCAGAACTTTAGGAACTTTATCCTAGAGCGCTCAGGAATCTTGCCCGCCACCTGCT
GTGC
ACTTCCTAGCGACTTTGTGCCCATTAAGTACCGCGAATGCCCTCCGCCGCTTTGGG
GCCA
CTGCTACCTTCTGCAGCTAGCCAACTACCTTGCCCTACCACTCTGACATAATGGAAG
ACGT
GAGCGGTGACGGTCTACTGGAGTGTCAGTGTGCTGCAACCTATGCACCCCGCAC
CGCTC
CCTGGTTTGCAATTCGCAGCTGCTTAACGAAAGTCAAATTATCGGTACCTTTGAGC
TGCA
GGGTCCCTCGCCTGACGAAAAGTCCGCGGCTCCGGGGTTGAACTCACTCCGGGG
CTGTG

GACGTCGGCTTACCTTCGCAAATTTGTACCTGAGGACTACCACGCCACGAGATT
AGGTT
CTACGAAGACCAATCCCGCCCGCAAATGCGGAGCTTACCGCCTGCGTCATTACC
CAGGG
CCACATTCTTGGCCAATTGCAAGCCATCAACAAAGCCCGCCAAGAGTTTCTGCTA
CGAAA
GGGACGGGGGGTTTACTTGGACCCCCAGTCCGGCGAGGAGCTCAACCCAATCCCC
CCGCC
GCCGCAGCCCTATCAGCAGCAGCCGCGGGCCCTTGCTTCCCAGGATGGCACCCAA
AAAGA
AGCTGCAGCTGCCGCCGCCACCCACGGACGAGGAGGAATACTGGGACAGTCAGG
CAGAGG
AGGTTTTGGACGAGGAGGAGGAGGACATGATGGAAGACTGGGAGAGCCTAGACG
AGGAAG
CTTCCGAGGTCGAAGAGGTGTCAGACGAAACACCGTCACCCTCGGTTCGCATTCCC
CTCGC
CGGCGCCCCAGAAATCGGCAACCGTTCCAGCATGGCTACAACCTCCGCTCCTCA
GGCGC
CGCCGGCACTGCCCGTTCGCCGACCCAACCGTAGATGGGACACCACTGGAACCAG
GGCCG
GTAAGTCCAAGCAGCCGCCGCCGTTAGCCCAAGAGCAACAACAGCGCCAAGGCT
ACCGCT
CATGGCGCGGGCACAAGAACGCCATAGTTGCTTGCTTGCAAGACTGTGGGGGCAA
CATCT
CCTTCGCCCCGCCGCTTCTTCTCTACCATCACGGCGTGGCCTTCCCCCGTAACATC
CTGC
ATTACTACCGTCATCTCTACAGCCATACTGCACCGGCGGCAGCGGCAGCGGCAG
CAACA
GCAGCGGCCACACAGAAGCAAAGGCGACCGGATAGCAAGACTCTGACAAAGCCC
AAGAAA

TCCACAGCGGCGGCAGCAGCAGGAGGAGGAGCGCTGCGTCTGGCGCCCAACGAA
CCCGTA
TCGACCCGCGAGCTTAGAAACAGGATTTTTCCCACTCTGTATGCTATATTTCAACA
GAGC
AGGGGCCAAGAACAAGAGCTGAAAATAAAAAACAGGTCTCTGCGATCCCTCACC
CGCAGC
TGCCTGTATCACAAAAGCGAAGATCAGCTTCGGCGCACGCTGGAAGACGCGGAG
GCTCTC
TTCAGTAAATACTGCGCGCTGACTCTTAAGGACTAGTTTCGCGCCCTTTCTCAAAT
TTAA
GCGCGAAAACACTACGTCATCTCCAGCGGCCACACCCGGCGCCAGCACCTGTCGTCA
GCGCC
ATTATGAGCAAGGAAATTCCCACGCCCTACATGTGGAGTTACCAGCCACAAATGG
GACTT
GCGGCTGGAGCTGCCCAAGACTACTCAACCCGAATAAACTACATGAGCGCGGGA
CCCCAC
ATGATATCCCGGGTCAACGGAATCCGCGCCACCGAAACCGAATTCTCTTGGAAC
AGGCG
GCTATTACCACCACACCTCGTAATAACCTTAATCCCCGTAGTTGGCCCGCTGCCCT
GGTG
TACCAGGAAAGTCCCGCTCCCACCACTGTGGTACTTCCCAGAGACGCCAGGCCG
AAGTT
CAGATGACTAACTCAGGGGCGCAGCTTGCGGGCGGCTTTCGTCACAGGGTGCGGT
CGCC
GGCAGGGTATAACTCACCTGACAATCAGAGGGCGAGGTATTCAGCTCAACGAC
GAGTCG
GTGAGCTCCTCGCTTGGTCTCCGTCCGGACGGGACATTCAGATCGGCGGGCGCCG
GCCGT
CCTTCATTCACGCCTCGTCAGGCAATCCTAACTCTGCAGACCTCGTCCTCTGAGCC
GCGC

TCTGGAGGCATTGGA ACTCTGCAATTTATTGAGGAGTTTGTGCCATCGGTCTACTT
TAAC
CCCTTCTCGGGACCTCCCGGCCACTATCCGGATCAATTTATTCCCTAACTTTGACGC
GGTA
AAGGACTCGGCGGACGGCTACGACTGAATGTTAAGTGGAGAGGCAGAGCAACTG
CGCCTG
AAACACCTGGTCCACTGTCGCCGCCACAAGTGCTTTGCCCGCGACTCCGGTGAGT
TTTGC
TACTTTGAATTGCCCGAGGATCATATCGAGGGCCCGGCACGGCGTCCGGCTTA
CCGCC
CAGGGAGAGCTTGCCCGTAGCCTGATTCGGGAGTTTACCCAGCGCCCCCTGCTAG
TTGAG
CGGGACAGGGGACCCTGTGTTCTCACTGTGATTTGCAACTGTCCTAACCTTGGATT
ACAT
CAAGATCTTTGTTGCCATCTCTGTGCTGAGTATAATAAATACAGAAATTAATAAT
ACTG
GGGCTCCTATCGCCATCCTGTAAACGCCACCGTCTTCACCCGCCCAAGCAAACCA
AGGCG
AACCTTACCTGGTACTTTTAACATCTCTCCCTCTGTGATTTACAACAGTTTCAACCC
AGA
CGGAGTGAGTCTACGAGAGAACCTCTCCGAGCTCAGCTACTCCATCAGAAAAAAC
ACCAC
CCTCCTTACCTGCCGGGAACGTACGAGTGCGTCACCGGCCGCTGCACCACACCTA
CCGCC
TGACCGTAAACCAGACTTTTTCCGGACAGACCTCAATAACTCTGTTTACCAGAAC
AGGAG
GTGAGCTTAGAAAACCCTTAGGGTATTAGGCCAAAGGCGCAGCTACTGTGGGGTT
TATGA
ACAATTCAAGCAACTCTACGGGCTATTCTAATTCAGGTTTCTCTAGAATCGGGGTT
GGGG

TTATTCTCTGTCTTGTGATTCTCTTTATTCTTATACTAACGCTTCTCTGCCTAAGGC
TCG
CCGCCTGCTGTGTGCACATTTGCATTTATTGTCAGCTTTTTAAACGCTGGGGTCGC
CACC
CAAGATGATTAGGTACATAATCCTAGGTTTACTCACCTTGCGTCAGCCCACGGTA
CCAC
CCAAAAGGTGGATTTTAAGGAGCCAGCCTGTAATGTTACATTCGCAGCTGAAGCT
AATGA
GTGCACCACTCTTATAAAAATGCACCACAGAACATGAAAAGCTGCTTATTCGCCAC
AAAAA
CAAAATTGGCAAGTATGCTGTTTATGCTATTTGGCAGCCAGGTGACACTACAGAG
TATAA
TGTTACAGTTTTCCAGGGTAAAAGTCATAAACTTTTATGTATACTTTTCCATTTTA
TGA
AATGTGCGACATTACCATGTACATGAGCAAACAGTATAAGTTGTGGCCCCACAA
AATTG
TGTGGAAAACACTGGCACTTTCTGCTGCACTGCTATGCTAATTACAGTCTCGCTT
TGGT
CTGTACCCTACTCTATATTAATAACAAAAGCAGACGCAGCTTTATTGAGGAAAAG
AAAAT
GCCTTAATTTACTAAGTTACAAAGCTAATGTCACCACTAACTGCTTTACTCGCTGC
TTGC
AAAACAAATTCAAAAAGTTAGCATTATAATTAGAATAGGATTTAAACCCCCGGT
CATTT
CCTGCTCAATACCATTCCCCTGAACAATTGACTCTATGTGGGATATGCTCCAGCGC
TACA
ACCTGAAGTCAGGCTTCCTGGATGTCAGCATCTGACTTTGGCCAGCACCTGTCCC
GCGG
ATTTGTTCCAGTCCAACACTACAGCGACCCACCCTAACAGAGATGACCAACACAACC
AACGC

GGCCGCCGCTACCGGACTTACATCTACCACAAATACACCCCAAGTTTCTGCCTTTG
TCAA
TAACTGGGATAACTTGGGCATGTGGTGGTTCTCCATAGCGCTTATGTTTGTATGCC
TTAT
TATTATGTGGCTCATCTGCTGCCTAAAGCGCAAACGCGCCCGACCACCCATCTATA
GTCC
CATCATTGTGCTACACCCAAACAATGATGGAATCCATAGATTGGACGGACTGAAA
CACAT
GTTCTTTTCTCTTACAGTATGATTAAATGAGACATGATTCCTCGAGTTTTTATATTA
CTG
ACCCTTGTTGCGCTTTTTTGTGCGTGCTCCACATTGGCTGCGGTTTCTCACATCGAA
GTA
GACTGCATTCCAGCCTTCACAGTCTATTTGCTTTACGGATTTGTCACCCTCACGCT
CATC
TGCAGCCTCATCACTGTGGTCATCGCCTTTATCCAGTGCATTGACTGGGTCTGTGT
GCGC
TTTGCATATCTCAGACACCATCCCCAGTACAGGGACAGGACTATAGCTGAGCTTC
TTAGA
ATTCTTTAATTATGAAATTTACTGTGACTTTTCTGCTGATTATTTGCACCCTATCTG
CGT
TTTGTCCCCGACCTCCAAGCCTCAAAGACATATATCATGCAGATTCACCTCGTATA
TGGA
ATATTCCAAGTTGCTACAATGAAAAAAGCGATCTTTCCGAAGCCTGGTTATATGC
AATCA
TCTCTGTTATGGTGTCTGCAGTACCATCTTAGCCCTAGCTATATATCCCTACCTTG
ACA
TTGGCTGGAACGCAATAGATGCCATGAACCACCCAACTTTCCCCGCGCCCGCTAT
GCTTC
CACTGCAACAAGTTGTTGCCGGCGGCTTTGTCCCAGCCAATCAGCCTCGCCCACCT
TCTC

CCACCCCCACTGAAATCAGCTACTTTAATCTAACAGGAGGAGATGACTGACACCC
TAGAT
CTAGA A ATGGACGGAATTATTACAGAGCAGCGCCTGCTAGAAAGACGCAGGGCA
GCGGCC
GAGCAACAGCGCATGAATCAAGAGCTCCAAGACATGGTAACTTGCACCAGTGCA
AAAGG
GGTATCTTTTGTCTGGTAAAGCAGGCCAAAGTCACCTACGACAGTAATACCACCG
GACAC
CGCCTTAGCTACAAGTTGCCAACCAAGCGTCAGA A ATTGGTGGTCATGGTGGGAG
AAAAG
CCCATTACCATAACTCAGCACTCGGTAGAAACCGAAGGCTGCATTCACCTCACCTT
GTCAA
GGACCTGAGGATCTCTGCACCCTTATTAAGACCCTGTGCGGTCTCAAAGATCTTAT
TCCC
TTTAACTAATAAAAAAAAAATAATAAGCATCACTTACTTAAAATCAGTTAGCAAA
TTTCT
GTCCAGTTTATTTCAGCAGCACCTCCTTGCCCTCCTCCCAGCTCTGGTATTGCAGCT
TCCT
CCTGGCTGCAAACCTTTCTCCACAATCTA 从 TGG AATGTCAGTTTCCTCCTGTTCT
GTCC
ATCCGCACCCACTATCTTCATGTTGTTGCAGATGAAGCGCGCAAGACCGTCTGAA
GATAC
CTTCAACCCCGTGTATCCATATGACACGGAAACCGGTCCTCCAACCTGTGCCTTTTC
TTAC
TCCTCCCTTTGTATCCCCAATGGGTTTCAAGAGAGTCCCCCTGGGGTACTCTCTT
TGCG
CCTATCCGAACCTCTAGTTACCTCCAATGGCATGCTTGCGCTCAAAATGGGCAAC
GGCCT
CTCTCTGGACGAGGCCGGCAACCTTACCTCCCAAATGTAACCACTGTGAGCCCA
CCTCT

CAAAAAACCAAGTCAAACATAAACCTGGAAATATCTGCACCCCTCACAGTTACC
TCAGA
AGCCCTAACTGTGGCTGCCGCCGCACCTCTAATGGTCGCGGGCAACACACTCACC
ATGCA
ATCACAGGCCCCGCTAACCGTGCACGACTCCAAACTTAGCATTGCCACCCAAGGA
CCCCCT
CACAGTGTGAGAAGGAAAGCTAGCCCTGCAAACATCAGGCCCCCTCACCACCACC
GATAG
CAGTACCCTTACTATCACTGCCTCACCCCTCTAACTACTGCCACTGGTAGCTTGG
GCAT
TGACTTGAAAGAGCCATTTATACACAAAATGGAAAAGTAGGACTAAAGTACGG
GGCTCC
TTTGCATGTAACAGACGACCTAAACACTTTGACCGTAGCAACTGGTCCAGGTGTG
ACTAT
TAATAATACTTCCTTGCAAACATAAGTTACTGGAGCCTTGGGTTTTGATTCACAAG
GCAA
TATGCAACTTAATGTAGCAGGAGGACTAAGGATTGATTCTCAAAACAGACGCCTT
ATACT
TGATGTTAGTTATCCGTTTGATGCTCAAAACCAACTAAATCTAAGACTAGGACAG
GGCCC
TCTTTTTATAAACTCAGCCCACAACCTGGATATTAACATAACAAAGGCCTTACT
TGTT
TACAGCTTCAAACAATTCCAAAAAGCTTGAGGTAACTAAGCACTGCCAAGGGG
TTGAT
GTTTGACGCTACAGCCATAGCCATTAATGCAGGAGATGGGCTTGAATTTGGTTCA
CCTAA
TGCACCAAACACAAATCCCCTCAAAACAAAAATTGGCCATGGCCTAGAATTTGAT
TCAAA
CAAGGCTATGGTTCCTAAACTAGGAACTGGCCTTAGTTTTGACAGCACAGGTGCC
ATTAC

AGTAGGAAACAAAAATAATGATAAGCTAACTTTGTGGACCACACCAGCTCCATCT
CCTAA
CTGTAGACTAAATGCAGAGAAAGATGCTAAACTCACTTTGGTCTTAACAAAATGT
GGCAG
TCAAATACTTGCTACAGTTTCAGTTTTGGCTGTAAAGGCAGTTTGGCTCCAATAT
CTGG
AACAGTTCAAAGTGCTCATCTTATTATAAGATTTGACGAAAATGGAGTGCTACTA
AACAA
TTCCTTCCTGGACCCAGAATATTGGAACCTTAGAAATGGAGATCTTACTGAAGGC
ACAGC
CTATACAAACGCTGTTGGATTTATGCCTAACCTATCAGCTTATCCAAAATCTCACG
GTAA
AACTGCCAAAAGTAACATTGTCAGTCAAGTTTACTTAAACGGAGACAAAATAAA
CCTGT
AACACTAACCATTAACACTAAACGGTACACAGGAAACAGGAGACACAACCTCCAAG
TGCATA
CTCTATGTCATTTTCATGGGACTGGTCTGGCCACAACACTACATTAATGAAATATTTG
CCAC
ATCCTCTTACACTTTTTTCATACATTGCCCAAGAATAAAGAATCGTTTGTGTTATGT
TTCA
ACGTGTTTATTTTTCAATTGCAGAAAATTTCAAGTCATTTTTCATTCAGTAGTATA
GCCC
CACCACCACATAGCTTATACAGATCACCGTACCTTAATCAAACCTCACAGAACCCT
AGTAT
TCAACCTGCCACCTCCCTCCCAACACACAGAGTACACAGTCCTTTCTCCCCGGCTG
GCCT
TAAAAAGCATCATATCATGGGTAACAGACATATTCTTAGGTGTTATATTCCACAC
GGTTT
CCTGTCGAGCCAAACGCTCATCAGTGATATTAATAAACTCCCCGGGCAGCTCACT
TAAGT

TCATGTCGCTGTCCAGCTGCTGAGCCACAGGCTGCTGTCCAACCTTGCGGTTGCTTA
ACGG
GCGGCGAAGGAGAAGTCCACGCCTACATGGGGGTAGAGTCATAATCGTGCATCA
GGATAG
GGCGGTGGTGCTGCAGCAGCGCGCAATAAACTGCTGCCGCCGCCGCTCCGTCTT
GCAGG
AATACAACATGGCAGTGGTCTCCTCAGCGATGATTCGCACCGCCCGCAGCATAAG
GCGCC
TTGTCCTCCGGGCACAGCAGCGCACCCCTGATCTCACTTAAATCAGCACAGTAACT
GCAGC
ACAGCACCACAATATTGTTCAAAATCCCACAGTGCAAGGCGCTGTATCCAAAGCT
CATGG
CGGGGACCACAGAACCCACGTGGCCATCATAACCACAAGCGCAGGTAGATTAAGT
GGCGAC
CCCTCATAAACACGCTGGACATAAACATTACCTCTTTTGGCATGTTGTAATTCACC
ACCT
CCCGGTACCATATAAACCTCTGATTAAACATGGCGCCATCCACCACCATCCTAAA
CCAGC
TGGCCAAAACCTGCCCGCCGGCTATACTGCAGGGAACCGGGACTGGAACAAT
GACAGT
GGAGAGCCCAGGACTCGTAACCATGGATCATCATGCTCGTCATGATATCAATGTT
GGCAC
AACACAGGCACACGTGCATACACTTCCTCAGGATTACAAGCTCCTCCCGCGTTAG
AACCA
TATCCCAGGGAACAACCCATTCTGAATCAGCGTAAATCCCACACTGCAGGGAAG
ACCTC
GCACGTAACCTCACGTTGTGCATTGTCAAAGTGTTACATTCGGGCAGCAGCGGATG
ATCCT
CCAGTATGGTAGCGCGGGTTTCTGTCTCAAAGGAGGTAGACGATCCCTACTGTA
CGGAG

TGCGCCGAGACAACCGAGATCGTGTTGGTCGTAGTGTCATGCCAAATGGAACGCC
GGACG
TAGTCATATTTCTGAAGCAAAACCAGGTGCGGGCGTGACAAACAGATCTGCGTC
TCCGG
TCTCGCCGCTTAGATCGCTCTGTGTAGTAGTTGTAGTATATCCACTCTCTCAAAGC
ATCC
AGGCGCCCCCTGGCTTCGGGTTCTATGTAAACTCCTTCATGCGCCGCTGCCCTGAT
AACA
TCCACCACCGCAGAATAAGCCACACCCAGCCAACCTACACATTCGTTCTGCGAGT
CACAC
ACGGGAGGAGCGGGAAGAGCTGGAAGAACCATGTTTTTTTTTTTATTCCAAAAGA
TTATC
CAAAACCTCAAAATGAAGATCTATTAAGTGAACGCGCTCCCCTCCGGTGGCGTGG
TCAA
CTCTACAGCCAAAGAACAGATAATGGCATTGTGAAGATGTTGCACAATGGCTTCC
AAAAG
GCAAACGGCCCTCACGTCCAAGTGGACGTAAAGGCTAAACCCTTCAGGGTGAATC
TCCTC
TATAAACATTCCAGCACCTTCAACCATGCCCAAATAATTCTCATCTCGCCACCTTC
TCAA
TATATCTCTAAGCAAATCCCGAATATTAAGTCCGGCCATTGTAAAAATCTGCTCCA
GAGC
GCCCTCCACCTTCAGCCTCAAGCAGCGAATCATGATTGCAAAAATTCAGGTTCTC
ACAG
ACCTGTATAAGATTCAAAGCGGAACATTAACAAAAATACCGCGATCCCGTAGGT
CCCTT
CGCAGGGCCAGCTGAACATAATCGTGCAGGTCTGCACGGACCAGCGGGCCACTT
CCCCG
CCAGGAACCTTGACAAAAGAACCACACTGATTATGACACGCATACTCGGAGCTA
TGCTA

ACCAGCGTAGCCCCGATGTAAGCTTTGTTGCATGGGCGGCGATATAAAAATGCAAG
GTGCT
GCTCAAAAAATCAGGCAAAGCCTCGCGCAAAAAAGAAAGCACATCGTAGTCATG
CTCATG
CAGATAAAGGCAGGTAAGCTCCGGAACCACCACAGAAAAGACACCATTTTTCTC
TCAA
CATGTCTGCGGGTTTCTGCATAAACACAAAATAAAATAACAAAAAACATTTAA
CATT
GAAGCCTGTCTTACAACAGGAAAAACAACCCTTATAAGCATAAGACGGACTACG
GCCATG
CCGGCGTGACCGTAAAAAACTGGTCACCGTGATTA AAAAGCACACCACGACAGCT
CCTCG
GTCATGTCCGGAGTCATAATGTAAGACTCGGTAAACACATCAGGTTGATTCATCG
GTCAG
TGCTAAAAAGCGACCGAAATAGCCCGGGGAATACATACCCGCAGGCGTAGAGA
CAACAT
TACAGCCCCATAGGAGGTATAACAAAATTAATAGGAGAGAAAAACACATAAAC
ACCTGA
AAAACCTCCTGCCTAGGCAAAATAGCACCTCCCGCTCCAGAACAACATACAGC
GCTTC
CACAGCGGCAGCCATAACAGTCAGCCTTACCAGTAAAAAAGAAAACCTATTTAA
AAAACA
CCACTCGACACGGCACCCAGCTCAATCAGTCACAGTGTAAAAAAGGGCCAAGTGC
AGAGCG
AGTATATATAGGACTAAAAATGACGTAACGGTTAAAGTCCACAAAAACACCC
AGAAAA
CCGCACGCGAACCTACGCCAGAAACGAAAGCCAAAAAACCACAACCTTCCTCA
AATCGT
CACTTCGGTTTTCCACGTTACGTCACCTCCCATTTTAATTAAGAAAACCTACAATT
CCCA

ACACATACAAGTTACTCCGCCCTAAAACCTACGTCACCCGCCCCGTTCCCACGCC
CGCG
CCACGTCACAACTCCACCCCTCATTATCATATTGGCTTCAATCCAAAATAAGGT
ATAT
TATTGATGATGATTACCCTGTTAT
SEQ ID NO: 50 是 OV1164 序列
CAGGGTAATCATCATCAATAATATACCTTATTTGGATTGAAGCCAATATGATAAT
GAGG
GGGTGGAGTTTGTGACGTGGCGCGGGGCGTGGGAACGGGGCGGGTGACGTAGTA
GTGTGG
CGGAAGTGTGATGTTGCAAGTGTGGCGGAACACATGTAAGCGACGGATGTGGCA
AAAGTG
ACGTTTTTGGTGTGCGCCGGTGTACACAGGAAGTGACAATTTTCGCGCGGTTTTAG
GCGG
ATGTTGTAGTAAATTTGGGCGTAACCGAGTAAGATTTGGCCATTTTCGCGGGAAA
ACTGA
ATAAGAGGAAGTGAAATCTGAATAATTTTGTGTTACTCATAGCGCGTAATATTTGT
CTAG
GGCCGGGATCTCTGCAGGAATTTGATATCAAGCTTATCGATACCGTCGAAACTTG
TTTAT
TGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAA
GCATT
TTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGT
CTG
GATCCGCTAGCGGCGCGCCGTTTCATCCGGACAAAGCCTGCGCGCGCCCCGCC
GCCAT
TGGCCGTACCGCCCCGCGCCGCCGCCCATCTCGCCCCTCGCCGCCGGGTCCGGC
GCGTT
AAAGCCAATAGGAACCGCCGCGTTGTTCCCGTCACGGCCGGGGCAGCCAATTGT
GGCGG

CGCTCGGCGGCTCGTGGCTCTTTCGCGGCAAAAAGGATTTGGCGCGTAAAAGTGG
CCGGG
ACTTTGCAGGCAGCGGCGGCCGGGGCGGAGCGGGATCGAGCCCTCGATGATAT
CAGATC
AAACGATATCACCGGTCGACTGAAAATGAGACATATTATCTGCCACGGAGGTGTT
ATTAC
CGAAGAAATGGCCGCCAGTCTTTTGGACCAGCTGATCGAAGAGGTAAGTGGCTGAT
AATCT
TCCACCTCCTAGCCATTTTGAACCACCTACCCTTCACGAACTGTATGATTTAGACG
TGAC
GGCCCCGAAGATCCCAACGAGGAGGCGGTTTCGCAGATTTTCCCGACTCTGTA
ATGTT
GGCGGTGCAGGAAGGGATTGACTTACTCACTTTTCCGCCGGCGCCCGTTCTCCG
GAGCC
GCCTCACCTTCCCGGCAGCCCGAGCAGCCGGAGCAGAGAGCCTTGGGTCCGGTT
TCTAT
GCCAAACCTTGTACCGGAGGTGATCGATCTTACCTGCCACGAGGCTGGCTTTCCA
CCCAG
TGACGACGAGGATGAAGAGGGTGAGGAGTTTGTGTTAGATTATGTGGAGCACCCC
GGGCA
CGGTTGCAGGTCTTGTCATTATCACCGGAGGAATACGGGGGACCCAGATATTATG
TGTTT
GCTTTGCTATATGAGGACCTGTGGCATGTTTGTCTACAGTAAGTGAAAATTATGGG
CAGT
GGGTGATAGAGTGGTGGGTTTGGTGTGGTAATTTTTTTTTTAATTTTACAGTTTTG
TGG
TTTAAAGAATTTTGTATTGTGATTTTTTTTAAAAGGTCCTGTGTCTGAACCTGAGCC
TGAG
CCCGAGCCAGAACCGGAGCCTGCAAGACCTACCCGCCGTCCTAAAATGGCGCCTG
CTATC

CTGAGACGCCCGACATCACCTGTGTCTAGAGAATGCAATAGTAGTACGGATAGCT
GTGAC
TCCGGTCCTTCTAACACACCTCCTGAGATACACCCGGTGGTCCCGCTGTGCCCCAT
TAAA
CCAGTTGCCGTGAGAGTTGGTGGGCGTCGCCAGGCTGTGGAATGTATCGAGGACT
TGCTT
AACGAGCCTGGGCAACCTTTGGACTTGAGCTGTAAACGCCCCAGGCCATAAGGTG
TAAAC
CTGTGATTGCGTGTGTGGTTAACGCCTTTGTTTGCTGAATGAGTTGATGTAAGTTT
AATA
AAGGGTGAGATAATGTTTAACTTGCATGGCGTGTTAAATGGGGCGGGGCTTAAAG
GGTAT
ATAATGCGCCGTGGGCTAATCTTGGTTACATCTGACCTCATGGAGGCTTGGGAGT
GTTTG
GAAGATTTTTCTGCTGTGCGTAACTTGCTGGAACAGAGCTCTAACAGTACCTCTTG
GTTT
TGGAGGTTTCTGTGGGGCTCATCCCAGGCAAAGTTAGTCTGCAGAATTAAGGAGG
ATTAC
AAGTGGGAATTTGAAGAGCTTTTGAATCCTGTGGTGAGCTGTTTGATTCTTTGAA
TCTG
GGTCACCAGGCGCTTTTCCAAGAGAAGGTCATCAAGACTTTGGATTTTTCCACACC
GGGG
CGCGCTGCGGCTGCTGTTGCTTTTTTGAGTTTTATAAAGGATAAATGGAGCGAAG
AAACC
CATCTGAGCGGGGGGTACCTGCTGGATTTTCTGGCCATGCATCTGTGGAGAGCGG
TTGTG
AGACACAAGAATCGCCTGCTACTGTTGTCTTCCGTCCGCCCGGCGATAATACCGA
CGGAG
GAGCAGCAGCAGCAGCAGGAGGAAGCCAGGCGGCGGCGGCAGGAGCAGAGCCC
ATGGAAC

CCGAGAGCCGGCCTGGACCCTCGGGAATGAATGTTGTACAGGTGGCTGAACTGTA
TCCAG
AACTGAGACGCATTTTGACAATTACAGAGGATGGGCAGGGGCTAAAGGGGGTAA
AGAGGG
AGCGGGGGGCTTGTGAGGCTACAGAGGAGGCTAGGAATCTAGCTTTTAGCTTAAT
GACCA
GACACCGTCCTGAGTGTATTACTTTTCAACAGATCAAGGATAATTGCGCTAATGA
GCTTG
ATCTGCTGGCGCAGAAGTATTCCATAGAGCAGCTGACCACTTACTGGCTGCAGCC
AGGGG
ATGATTTTGAGGAGGCTATTAGGGTATATGCAAAGGTGGCACTTAGGCCAGATTG
CAAGT
ACAAGATCAGCAAACCTGTAAATATCAGGAATTGTTGCTACATTTCTGGGAACGG
GGCCG
AGGTGGAGATAGATACGGAGGATAGGGTGGCCTTTAGATGTAGCATGATAAATAT
GTGGC
CGGGGGTGCTTGGCATGGACGGGGTGGTTATTATGAATGTAAGGTTTACTGGCCC
CAATT
TTAGCGGTACGGTTTTCTGGCCAATACCAACCTTATCCTACACGGTGTAAGCTTC
TATG
GGTTAACAATACCTGTGTGGAAGCCTGGACCGATGTAAGGGTTCGGGGCTGTGC
CTTTT
ACTGCTGCTGGAAGGGGGTGGTGTGTCGCCCAAAGCAGGGCTTCAATTAAGAA
ATGCC
TCTTTGAAAGGTGTACCTTGGGTATCCTGTCTGAGGGTAACTCCAGGGTGCGCCAC
AATG
TGGCCTCCGACTGTGGTTGCTTCATGCTAGTGAAAAGCGTGGCTGTGATTAAGCAT
AACA
TGGTATGTGGCAACTGCGAGGACAGGGCCTCTCAGATGCTGACCTGCTCGGACGG
CAACT

GTCACCTGCTGAAGACCATTACGTAGCCAGCCACTCTCGCAAGGCCTGGCCAGT
GTTTG
AGCATAACATACTGACCCGCTGTTCCCTTGCATTTGGGTAACAGGAGGGGGGTGTT
CCTAC
CTTACCAATGCAATTTGAGTCACACTAAGATATTGCTTGAGCCCGAGAGCATGTC
CAAGG
TGAACCTGAACGGGGTGTGGACATGACCATGAAGATCTGGAAGGTGCTGAGGTA
CGATG
AGACCCGCACCAGGTGCAGACCCTGCGAGTGTGGCGGTAAACATATTAGGAACC
AGCCTG
TGATGCTGGATGTGACCGAGGAGCTGAGGCCGATCACTTGGTGCTGGCCTGCAC
CCGCG
CTGAGTTTGGCTCTAGCGATGAAGATACAGATTGAGGTACTGAAATGTGTGGGCG
TGGCT
TAAGGGTGGGAAAGAATATATAAGGTGGGGTCTTATGTAGTTTTGTATCTGTTTT
GCAG
CAGCCGCCGCCCATGAGCACCAACTCGTTTGATGGAAGCATTGTGAGCTCATA
TTTGA
CAACGCGCATGCCCCATGGGCCGGGGTTCGTCAGAATGTGATGGGCTCCAGCAT
TGATG
GTCGCCCCGTCCTGCCCGCAAACCTCTACTACCTTGACCTACGAGACCGTGTCTGGA
ACGC
CGTTGGAGACTGCAGCCTCCGCCCGCTTCAGCCGCTGCAGCCACCGCCCGCGG
GATTG
TGA CTGACTTTGCTTTCCTGAGCCCGCTTGCAAGCAGTGCAGCTTCCCGTTCATCC
GCCC
GCGATGACAAGTTGACGGCTCTTTTGGCACAATTGGATTCTTTGACCCGGGAACCT
AATG
TCGTTTCTCAGCAGCTGTTGGATCTGCGCCAGCAGGTTTCTGCCCTGAAGGCTTCC
TCCC

CTCCAATGCGGTTTAAAACATAAATAAAAAACCAGACTCTGTTTGGATTGGAT
CAAGC
AAGTGTCTTGCTGTCTTTATTTAGGGGTTTTGCGCGCGCGGTAGGCCCGGGACCAG
CGGT
CTCGGTCGTTGAGGGTCCTGTGTATTTTTTCCAGGACGTGGTAAAGGTGACTCTGG
ATGT
TCAGATACATGGGCATAAGCCCGTCTCTGGGGTGGAGGTAGCACCCTGCAGAGC
TTCAT
GCTGCGGGGTGGTGTGTAGATGATCCAGTCGTAGCAGGAGCGCTGGGCGTGGTG
CCTAA
AAATGTCTTTCAGTAGCAAGCTGATTGCCAGGGGCAGGCCCTTGGTGTAAGTGTT
TACAA
AGCGGTTAAGCTGGGATGGGTGCATACGTGGGGATATGAGATGCATCTTGGACTG
TATTT
TTAGGTTGGCTATGTTCCCAGCCATATCCCTCCGGGGATTCATGTTGTGCAGAACC
ACCA
GCACAGTGTATCCGGTGCACCTTGGGAAATTTGTCATGTAGCTTAGAAGGAAATGC
GTGGA
AGA ACTTGGAGACGCCCTTGTGACCTCCAAGATTTTCCATGCATTCGTCCATAATG
ATGG
CAATGGGCCACGGGCGGCGGCCTGGGCGAAGATATTTCTGGGATCACTAACGTC
ATAGT
TGTGTTCCAGGATGAGATCGTCATAGGCCATTTTTACAAAGCGCGGGCGGAGGGT
GCCAG
ACTGCGGTATAATGGTTCCATCCGGCCCAGGGGCGTAGTTACCCTCACAGATTTG
CATTT
CCCACGCTTTGAGTTCAGATGGGGGGATCATGTCTACCTGCGGGGCGATGAAGAA
AACGG
TTTCCGGGGTAGGGGAGATCAGCTGGGAAGAAAGCAGGTTCCCTGAGCAGCTGCG
ACTTAC

CGCAGCCGGTGGGCCCCTAAATCACACCTATTACCGGGTGCAACTGGTAGTTAAG
AGAGC
TGCAGCTGCCGTCATCCCTGAGCAGGGGGGCCACTTCGTTAAGCATGTCCCTGAC
TCGCA
TGTTTTCCCTGACCAAATCCGCCAGAAGGCGCTCGCCGCCAGCGATAGCAGTTC
TTGCA
AGGAAGCAAAGTTTTTCAACGGTTTGAGACCGTCCGCCGTAGGCATGCTTTTGAG
CGTTT
GACCAAGCAGTTCAGGCGGTCCCACAGCTCGGTACCTGCTCTACGGCATCTCG
ATCCA
GCATATCTCCTCGTTTCGCGGGTGGGGCGGCTTTCGCTGTACGGCAGTAGTCGGT
GCTC
GTCCAGACGGGCCAGGGTCATGTCTTCCACGGGCGCAGGGTCCTCGTCAGCGTA
GTCTG
GGTACGGTGAAGGGGTGCGCTCCGGGCTGCGCGCTGGCCAGGGTGCCTTGAGG
CTGGT
CCTGCTGGTGCTGAAGCGCTGCCGGTCTTCGCCCTGCGCGTCGGCCAGGTAGCATT
TGAC
CATGGTGTCATAGTCCAGCCCCTCCGCGCGTGGCCCTTGGCGCGCAGCTTGCCCT
TGGA
GGAGGCGCCGCACGAGGGGCAGTGCAGACTTTTGAGGGCGTAGAGCTTGGGGCGC
GAGAAA
TACCGATTCCGGGGAGTAGGCATCCGCGCCGCAGGCCCCGCAGACGGTCTCGCAT
TCCAC
GAGCCAGGTGAGCTCTGGCCGTTTCGGGGTCAAAAACCAGGTTTCCCCATGCTTT
TTGAT
GCGTTTCTTACCTCTGGTTTCCATGAGCCGGTGTCCACGCTCGGTGACGAAAAGGC
TGTC
CGTGTCCCCGTATACAGACTTGAGAGGCCTGTCCCTCGAGCGGTGTTCCGCGGTCTC
CCTC

GTATAGAAACTCGGACCACTCTGAGACAAAGGCTCGCGTCCAGGCCAGCACGAA
GGAGGC
TAAGTGGGAGGGGTAGCGGTTCGTTGTCCACTAGGGGGTCCACTCGCTCCAGGGTG
TGAAG
ACACATGTCGCCCTCTTCGGCATCAAGGAAGGTGATTGGTTTGTAGGTGTAGGCC
ACGTG
ACCGGGTGTTCCTGAAGGGGGGCTATAAAAGGGGGTGGGGGCGCGTTCGTCTCTCA
CTCTC
TTCCGCATCGCTGTCTGCGAGGGCCAGCTGTTGGGGTGAGTACTCCCTCTGAAAA
GCGGG
CATGACTTCTGCGCTAAGATTGTCAGTTTCCAAAAACGAGGAGGATTTGATATTC
ACCTG
GCCCGCGGTGATGCCTTTGAGGGTGGCCGCATCCATCTGGTCAGAAAAGACAATC
TTTTT
GTTGTCAAGCTTGGTGGCAAACGACCCGTAGAGGGCGTTGGACAGCAACTTGGCG
ATGGA
GCGCAGGGTTTGGTTTTTGTGCGATCGGCGCGCTCCTTGGCCGCGATGTTAGCT
GCAC
GTATTCGCGCGCAACGCACCGCCATTCGGGAAAGACGGTGGTGCCTCGTCGGGC
ACCAG
GTGCACGCGCCAACCGCGTGTGTCAGGGTGACAAGGTCAACGCTGGTGGCTACC
TCTCC
GCGTAGGCGCTCGTTGGTCCAGCAGAGGCGGCCGCCCTTGC GCGAGCAGAATGGC
GGTAG
GGGGTCTAGCTGCGTCTCGTCCGGGGGTCTGCGTCCACGGTAAAGACCCCGGGC
AGCAG
GCGCGGTCTGAAGTAGTCTATCTTGCATCCTTGCAAGTCTAGCGCCTGCTGCCATG
CGCG
GGCGGCAAGCGCGCTCGTATGGGTTGAGTGGGGGACCCCATGGCATGGGGTG
GGTGAG

CGCGGAGGCGTACATGCCGCAAATGTCGTAAACGTAGAGGGGCTCTCTGAGTATT
CCAAG
ATATGTAGGGTAGCATCTTCCACCGCGGATGCTGGCGCGCACGTAATCGTATAGT
TCGTG
CGAGGGAGCGAGGAGGTCTGGGACCGAGGTTGCTACGGGCGGGCTGCTCTGCTCG
GAAGAC
TATCTGCCTGAAGATGGCATGTGAGTTGGATGATATGGTTGGACGCTGGAAGACG
TTGAA
GCTGGCGTCTGTGAGACCTACCGCGTCACGCACGAAGGAGGCGTAGGAGTCGCGC
AGCTT
GTTGACCAGCTCGGCGGTGACCTGCACGTCTAGGGCGCAGTAGTCCAGGGTTTCC
TTGAT
GATGTCATACTTATCCTGTCCCTTTTTTTTCCACAGCTCGCGGTTGAGGACAACT
CTTC
GCGGTCTTTCCAGTACTCTTGGATCGGAAACCCGTCGGCCTCCGAACGGTAAGAG
CCTAG
CATGTAGAACTGGTTGACGGCCTGGTAGGCGCAGCATCCCTTTTCTACGGGTAGC
GCGTA
TGCCTGCGCGCCTTCCGGAGCGAGGTGTGGGTGAGCGCAAAGGTGTCCCTGACC
ATGAC
TTTGAGGTACTGGTATTTGAAGTCAGTGTCTCGTCGCATCCGCCCTGCTCCCAGAGCA
AAAA
GTCCGTGCGCTTTTTGGAACGCGGATTTGGCAGGGCGAAGGTGACATCGTTGAAG
AGTAT
CTTTCCCGCGGAGGCATAAAGTTGCGTGTGATGCGGAAGGGTCCCGGCACCTCG
GAACG
GTTGTTAATTACCTGGGCGGCGAGCACGATCTCGTCAAAGCCGTTGATGTTGTGG
CCCAC
AATGTAAAGTTCCAAGAAGCGCGGGATGCCCTTGATGGAAGGCAATTTTTTAAGT
TCCTC

GTAGGTGAGCTCTTCAGGGGAGCTGAGCCCGTGCTCTGAAAGGGCCCAGTCTGCA
AGATG
AGGGTTGGAAGCGACGAATGAGCTCCACAGGTCACGGGCCATTAGCATTTGCAGG
TGGTC
GCGAAAGGTCCTAAACTGGCGACCTATGGCCATTTTTTCTGGGGTGATGCAGTAG
AAGGT
AAGCGGGTCTTGTTCCCAGCGGTCCCATCCAAGGTTGCGGGCTAGGTCTCGCGCG
GCAGT
CACTAGAGGCTCATCTCCGCCGAAC TTCATGACCAGCATGAAGGGCACGAGCTGC
TTCCC
AAAGGCCCCCATCCAAGTATAGGTCTCTACATCGTAGGTGACAAAGAGACGCTCG
GTGCG
AGGATGCGAGCCGATCGGGAAGAACTGGATCTCCCGCCACCAATTGGAGGAGTG
GCTATT
GATGTGGTGAAAGTAGAAGTCCCTGCGACGGGCCGAACACTCGTGCTGGCTTTTG
TAAAA
ACGTGCGCAGTACTGGCAGCGGTGCACGGGCTGTACATCCTGCACGAGGTTGACC
TGACG
ACCGCGCACAAGGAAGCAGAGTGGGAATTTGAGCCCCTCGCCTGGCGGGTTTGGC
TGGTG
GTCTTCTACTTCGGCTGCTTGTCCTTGACCGTCTGGCTGCTCGAGGGGAGTTACGG
TGGA
TCGGACCACCACGCCGCGGAGCCCAAAGTCCAGATGTCCGCGCGCGGCGGTCCG
AGCTT
GATGACAACATCGCGCAGATGGGAGCTGTCCATGGTCTGGAGCTCCCGCGGCGTC
AGGTC
AGGCGGGAGCTCCTGCAGGTTTACCTCGCATAGACGGGTCAGGGCGCGGGCTAGA
TCCAG
GTGATACCTAATTTCCAGGGGCTGGTTGGTGGCGGCGTCGATGGCTTGCAAGAGG
CCGCA

TCCCCGCGGCGCGACTACGGTACCGCGCGGCGGGCGGTGGGCCGCGGGGGTGTCC
TTGGA
TGATGCATCTAAAAGCGGTGACGCGGGCGAGCCCCGGAGGTAGGGGGGGCTCC
GGACCC
GCCGGGAGAGGGGGCAGGGGCACGTCGGCGCCGCGCGGGCAGGAGCTGGTGC
TGCGCG
CGTAGGTTGCTGGCGAACGCGACGACGCGGCGGTTGATCTCCTGAATCTGGCGCC
TCTGC
GTGAAGACGACGGGCCCGGTGAGCTTGAGCCTGAAAGAGAGTTCGACAGAATCA
ATTTCG
GTGTCGTTGACGGCGGCCTGGCGCAAATCTCCTGCACGTCTCCTGAGTTGTCTTG
ATAG
GCGATCTCGGCCATGAACTGCTCGATCTCTTCTCCTGGAGATCTCCGCGTCCGGC
TCGC
TCCACGGTGGCGGCGAGGTCGTTGAAATGCGGGCCATGAGCTGCGAGAAGGCG
TTGAGG
CCTCCCTCGTTCCAGACGCGGCTGTAGACCACGCCCCCTTCGGCATCGCGGGCGC
GCATG
ACCACCTGCGCGAGATTGAGCTCCACGTGCCGGGCGAAGACGGCGTAGTTTCGCA
GGCGC
TGAAAGAGGTAGTTGAGGGTGGTGGCGGTGTGTTCTGCCACGAAGAAGTACATAA
CCCAG
CGTCGCAACGTGGATTCGTTGATATCCCCCAAGGCCTCAAGGCGCTCCATGGCCT
CGTAG
AAGTCCACGGCGAAGTTGAAAACTGGGAGTTGCGCGCCGACACGGTTAACTCCT
CCTCC
AGAAGACGGATGAGCTCGGCGACAGTGTGCGGCACCTCGCGCTCAAAGGCTACA
GGGGCC
TCTTCTTCTTCAATCTCCTCTTCCATAAGGGCCTCCCCTTCTTCTTCTTCTGGCG
GC

GGTGGGGGAGGGGGGACACGGCGGCGACGACGGCGCACCGGGAGGCGGTTCGAC
AAAGCGC
TCGATCATCTCCCCGCGGCGACGGCGCATGGTCTCGGTGACGGCGCGGCCGTTCT
CGCGG
GGGCGCAGTTGGAAGACGCCGCCCGTCA TGTCCCGTTATGGGTTGGCGGGGGG
TGCCA
TGCGGCAGGGATACGGCGCTAACGATGCATCTCAACAATTGTTGTGTAGGTACTC
CGCCG
CCGAGGGACCTGAGCGAGTCCGCATCGACCGGATCGGAAAACCTCTCGAGAAAG
GCGTCT
AACCAGTCACAGTCGCAAGGTAGGCTGAGCACCGTGGCGGGCGGCAGCGGGCGG
CGGTCG
GGGTTGTTTCTGGCGGAGGTGCTGCTGATGATGTAATTAAAGTAGGCGGTCTTGA
GACGG
CGGATGGTCGACAGAAGCACCATGTCCTTGGGTCCGGCCTGCTGAATGCGCAGGC
GGTCG
GCCATGCCCCAGGCTTCGTTTTGACATCGGCGCAGGTCTTTGTAGTAGTCTTGCAT
GAGC
CTTTCTACCGGCACTTCTTCTTCCTTCCTTGTCTGTCCTGCATCTCTTGCATCTATCG
CT
GCGGCGGCGGCGGAGTTTGGCCGTAGGTGGCGCCCTCTTCCTCCCATGCGTGTGA
CCCCG
AAGCCCCTCATCGGCTGAAGCAGGGCTAGGTTCGGCGACAACGCGCTCGGCTAATA
TGGCC
TGCTGCACCTGCGTGAGGGTAGACTGGAAGTCATCCATGTCCACAAAGCGGTGGT
ATGCG
CCCGTGTTGATGGTGTAAGTGCAGTTGGCCATAACGGACCAGTTAACGGTCTGGT
GACCC
GGCTGCGAGAGCTCGGTGTACCTGAGACGCGAGTAAGCCCTCGAGTCAAATACGT
AGTCG

TTGCAAGTCCGCACCAGGTA CTGGTATCCCACCAAAAAGTGCGGCGGCGGCTGGC
GGTAG
AGGGGCCAGCGTAGGGTGGCCGGGGCTCCGGGGGCGAGATCTTCCAACATAAGG
CGATGA
TATCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGC
GCGGA
AAGTCGCGGACGCGGTTCCAGATGTTGCGCAGCGGCAAAAAGTGCTCCATGGTCCG
GGACG
CTCTGGCCGGTCAGGCGCGCAATCGTTGACGCTCTAGACCGTGCAAAAGGAGA
GCCTG
TAAGCGGGCACTCTTCCGTGGTCTGGTGGATAAATTCGCAAGGGTATCATGGCGG
ACGAC
CGGGGTTTCGAGCCCCGTATCCGGCCGTCCGCCGTGATCCATGCGGTTACCGCCCG
CGTGT
CGAACCCAGGTGTGCGACGTCAGACAACGGGGGAGTGCTCCTTTTGGCTTCCTTC
CAGGC
GCGGCGGCTGCTGCGCTAGCTTTTTTGGCCACTGGCCGCGCGCAGCGTAAGCGGT
TAGGC
TGGAAGCGAAAGCATTAAAGTGGCTCGCTCCCTGTAGCCGGAGGGTTATTTTCCA
AGGGT
TGAGTCGCGGGACCCCCGGTTCGAGTCTCGGACCGGCCGACTGCGGCGAACGGG
GGTTT
GCCTCCCCGTATGCAAGACCCCGCTTGCAAATTCCTCCGAAACAGGGACGAGC
CCCTT
TTTTGCTTTTCCCAGATGCATCCGGTGTGCGGCAGATGCGCCCCCTCCTCAGCA
GCGG
CAAGAGCAAGAGCAGCGGCAGACATGCAGGGCACCTCCCCTCCTCCTACCGCGT
CAGGA
GGGGCGACATCCGCGGTTGACGCGGCAGCAGATGGTGATTACGAACCCCCGCGG
CGCCGG

GCCCGGCACTACCTGGACTTGGAGGAGGGCGAGGGCCTGGCGCGGCTAGGAGCG
CCCTCT
CCTGAGCGGTACCCAAGGGTGCAGCTGAAGCGTGATACGCGTGAGGCGTACGTGC
CGCGG
CAGAACCTGTTTCGCGACCGCGAGGGAGAGGAGCCCGAGGAGATGCGGGATCGA
AAGTTC
CACGCAGGGCGCGAGCTGCGGCATGGCCTGAATCGCGAGCGGTTGCTGCGCGAG
GAGGAC
TTTGAGCCCGACGCGCGAACCGGGATTAGTCCCGCGCGCGCACACGTGGCGGCCG
CCGAC
CTGGTAACCGCATAACGAGCAGACGGTGAACCAGGAGATTAACCTTTCAAAAAGCT
TTAAC
AACCACGTGCGTACGCTTGTGGCGCGGAGGAGGTGGCTATAGGACTGATGCATC
TGTGG
GACTTTGTAAGCGCGCTGGAGCAAAACCCAAATAGCAAGCCGCTCATGGCGCAGC
TGTTT
CTTATAGTGCAGCACAGCAGGGACAACGAGGCATTCAGGGATGCGCTGCTAAAC
ATAGTA
GAGCCCGAGGGCCGCTGGCTGCTCGATTTGATAAACATCCTGCAGAGCATAGTGG
TGCAG
GAGCGCAGCTTGAGCCTGGCTGACAAGGTGGCCGCCATCAACTATTCCATGCTTA
GCCTG
GGCAAGTTTTACGCCCGCAAGATATACCATACCCCTTACGTTCCCATAGACAAGG
AGGTA
AAGATCGAGGGGTTCTACATGCGCATGGCGCTGAAGGTGCTTACCTTGAGCGACG
ACCTG
GGCGTTTATCGCAACGAGCGCATCCACAAGGCCGTGAGCGTGAGCCGGCGGCGC
GAGCTC
AGCGACCGCGAGCTGATGCACAGCCTGCAAAGGGCCCTGGCTGGCACGGGCAGC
GGCGAT

AGAGAGGCCGAGTCCTACTTTGACGCGGGCGCTGACCTGCGCTGGGCCCAAGCC
GACGC
GCCCTGGAGGCAGCTGGGGCCGGACCTGGGCTGGCGGTGGCACCCGCGCGCT
GGCAAC
GTCGGCGGCGTGGAGGAATATGACGAGGACGATGAGTACGAGCCAGAGGACGGC
GAGTAC
TAAGCGGTGATGTTTCTGATCAGATGATGCAAGACGCAACGGACCCGGCGGTGCG
GGCGG
CGCTGCAGAGCCAGCCGTCCGGCCTTAACTCCACGGACGACTGGCGCCAGGTCAT
GGACC
GCATCATGTCGCTGACTGCGCGCAATCCTGACGCGTTCCGGCAGCAGCCGCAGGC
CAACC
GGCTCTCCGCAATTCTGGAAGCGGTGGTCCCGGCGCGCGCAAACCCACGCACGA
GAAGG
TGCTGGCGATCGTAAACGCGCTGGCCGAAAACAGGGCCATCCGGCCCGACGAGG
CCGGCC
TGGTCTACGACGCGCTGCTTCAGCGCGTGGCTCGTTACAACAGCGGCAACGTGCA
GACCA
ACCTGGACCGGCTGGTGGGGGATGTGCGCGAGGCCGTGGCGCAGCGTGAGCGCG
CGCAGC
AGCAGGGCAACCTGGGCTCCATGGTTGCACTAAACGCCTTCCTGAGTACACAGCC
CGCCA
ACGTGCCGCGGGGACAGGAGGACTACACCAACTTTGTGAGCGCACTGCGGCTAAT
GGTGA
CTGAGACACCGCAAAGTGAGGTGTACCAGTCTGGGCCAGACTATTTTTTCCAGAC
CAGTA
GACAAGGCCTGCAGACCGTAAACCTGAGCCAGGCTTTCAAAACTTGCAGGGGCT
GTGGG
GGGTGCGGGCTCCCACAGGCGACCGCGCGACCGTGTCTAGCTTGCTGACGCCAA
CTCGC

GCCTGTTGCTGCTGCTAATAGCGCCCTTACGGACAGTGGCAGCGTGTCCCGGGA
CACAT
ACCTAGGTCACTTGCTGACACTGTACCGCGAGGCCATAGGTCAGGCGCATGTGGA
CGAGC
ATACTTTCCAGGAGATTACAAGTGTGACGCCGCGCGCTGGGGCAGGAGGACACGG
GCAGCC
TGGAGGCAACCCTAAACTACCTGCTGACCAACCGGCGGCAGAAGATCCCCTCGTT
GCACA
GTTTAAACAGCGAGGAGGAGCGCATTTCGCGCTACGTGCAGCAGAGCGTGAGCCT
TAACC
TGATGCGCGACGGGGTAACGCCAGCGTGGCGCTGGACATGACCGCGCGCAACA
TGGAAC
CGGGCATGTATGCCTCAAACCGGCCGTTTATCAACCGCCTAATGGACTACTTGCA
TGCAT
CGCG
CGGCCCGCGTGAACCCCGAGTATTTACCAATGCCATCTTGAACCCGCACTGGCT
ACCGC
CCCCTGGTTTCTACACCGGGGGATTTCGAGGTGCCCGAGGGTAACGATGGATTCT
CTGGG
ACGACATAGACGACAGCGTGTTCCTCCCGCAACCGCAGACCCTGCTAGAGTTGCA
ACAGC
GCGAGCAGGCAGAGGCGGCGCTGCGAAAGGAAAGCTTCCGCAGGCCAAGCAGCT
TGTCCG
ATCTAGGCGCTGCGGCCCGCGGTCAGATGCTAGTAGCCCATTTCCAAGCTTGAT
AGGGT
CTTTACCAGCACTCGCACCACCCGCCCGCGCCTGCTGGGCGAGGAGGAGTACCT
AAACA
ACTCGCTGCTGCAGCCGCGAGCGCGAAAAAACCTGCCTCCGGCATTTCCTCAACA
CGGGA
TAGAGAGCCTAGTGGACAAGATGAGTAGATGGAAGACGTACGCGCAGGAGCACA
GGGACG

TGCCAGGCCCGCGCCCGCCACCCGTCGTCAAAGGCACGACCGTCAGCGGGGTCT
GGTGT
GGGAGGACGATGACTCGGCAGACGACAGCAGCGTCCTGGATTTGGGAGGGAGTG
GCAACC
CGTTTGCGCACCTTCGCCCCAGGCTGGGGAGAATGTTTTAAAAAAAAAAAAAGCAT
GATGC
AAAATAAAAACTCACCAAGGCCATGGCACCGAGCGTTGGTTTTCTTGTATTCCC
CTTAG
TATGCGGCGCGCGGCGATGTATGAGGAAGGTCCTCCTCCCTCCTACGAGAGTGTG
GTGAG
CGCGGCGCCAGTGGCGGCGGCGCTGGGTTCTCCCTTCGATGCTCCCCTGGACCCG
CCGTT
TGTGCCTCCGCGGTACCTGCGGCCTACCGGGGGGAGAAACAGCATCCGTTACTCT
GAGTT
GGCACCCCTATTCGACACCACCCGTGTGTACCTGGTGGACAACAAGTCAACGGAT
GTGGC
ATCCCTGAACTACCAGAACGACCACAGCAACTTTCTGACCACGGTCATTCAAAC
AATGA
CTACAGCCCGGGGAGGCAAGCACACAGACCATCAATCTTGACGACCGGTCGCA
CTGGGG
CGGCGACCTGAAAACCATCCTGCATACCAACATGCCAAATGTGAACGAGTTCATG
TTTAC
CAATAAGTTTAAGGCGCGGGTGATGGTGTGCGGCTTGCCCTACTAAGGACAATCAG
GTGGA
GCTGAAATACGAGTGGGTGGAGTTCACGCTGCCCGAGGGCAACTACTCCGAGACC
ATGAC
CATAGACCTTATGAACAACGCGATCGTGGAGCACTACTTGAAAGTGGGCAGACAG
AACGG
GGTTCTGGAAAGCGACATCGGGGTAAAGTTTGACACCCGCAACTTCAGACTGGGG
TTTGA

CCCCGTCACTGGTCTTGTCATGCCTGGGGTATATACAAACGAAGCCTTCCATCCAG
ACAT
CATTTTGCTGCCAGGATGCGGGGTGGACTTCACCCACAGCCGCCTGAGCAACTTG
TTGGG
CATCCGCAAGCGGCAACCCTTCCAGGAGGGCTTTAGGATCACCTACGATGATCTG
GAGGG
TGGTAACATTCCCGCACTGTTGGATGTGGACGCCTACCAGGCGAGCTTGAAAGAT
GACAC
CGAACAGGGCGGGGGTGGCGCAGGCGGCAGCAACAGCAGTGGCAGCGGCGCGG
AAGAGAA
TCCAACGCGGCAGCCGCGGCAATGCAGCCGGTGGAGGACATGAACGATCATGC
ATTCG
CGGCGACACCTTTGCCACACGGGCTGAGGAGAAGCGCGCTGAGGCCGAAGCAGC
GGCCGA
AGCTGCCGCCCCCGCTGCGCAACCCGAGGTCGAGAAGCCTCAGAAGAAACCGGT
GATCAA
ACCCCTGACAGAGGACAGCAAGAAACGCAGTTACAACCTAATAAGCAATGACAG
CACCTT
CACCCAGTACCGCAGCTGGTACCTTGCATACA ACTACGGCGACCCTCAGACCGGA
ATCCG
CTCATGGACCCTGCTTTGCACTCCTGACGTAACCTGCGGCTCGGAGCAGGTCTACT
GGTC
GTTGCCAGACATGATGCAAGACCCCGTGACCTTCCGCTCCACGCGCCAGATCAGC
AACTT
TCCGGTGGTGGGCGCCGAGCTGTTGCCCGTGCACTCCAAGAGCTTCTACAACGAC
CAGGC
CGTCTACTCCCAACTCATCCGCCAGTTTACCTCTCTGACCCACGTGTTCAATCGCT
TTCC
CGAGAACCAGATTTTGGCGCGCCCGCCAGCCCCACCATCACCACCGTCAGTGAA
AACGT

TCCTGCTCTCACAGATCACGGGACGCTACCGCTGCGCAACAGCATCGGAGGAGTC
CAGCG
AGTGACCATTACTGACGCCAGACGCCGCACCTGCCCTACGTTTACAAGGCCCTG
GGCAT
AGTCTCGCCGCGCGTCCTATCGAGCCGCACTTTTTGAGCAAGCATGTCCATCCTTA
TATC
GCCCAGCAATAACACAGGCTGGGGCCTGCGCTTCCCAAGCAAGATGTTTGGCGGG
GCCAA
GAAGCGCTCCGACCAACACCCAGTGCGCGTGCGCGGGCACTACCGCGCGCCCTGG
GGCGC
GCACAAACGCGGCCGCACTGGGGCGACCACCGTCGATGACGCCATCGACGCGGT
GGTGGA
GGAGGCGCGCAACTACACGCCACGCCGCCACCAGTGTCCACAGTGGACGCGGC
CATTCA
GACCGTGGTGCGCGGAGCCCGGCGCTATGCTAAAATGAAGAGACGGCGGAGGCG
CGTAGC
ACGTCGCCACCGCCGCCGACCCGGCACTGCCGCCAACGCGCGGGCGGGCCCTG
CTTAA
CCGCGCACGTCGCACCGGCCGACGGGCGGCCATGCGGGCCGCTCGAAGGCTGGC
CGCGGG
ATTGTCACTGTGCCCCCAGGTCCAGGCGACGAGCGGCCGCCGACGAGCCGCG
GCCAT
TAGTGCTATGACTCAGGGTCGCAGGGGCAACGTGTATTGGGTGCGCGACTCGGTT
AGCGG
CCTGCGCGTGCCCGTGCGCACCCGCCCCCGCGCAACTAGATTGCAAGAAAAAAC
TACTT
AGACTCGTACTGTTGTATGTATCCAGCGGCGGGCGGCGCGCAACGAAGCTATGTCC
AAGCG
CAAATCAAAGAAGAGATGCTCCAGGTCATCGCGCCGGAGATCTATGGCCCCCG
AAGAA

GGAAGAGCAGGATTACAAGCCCCGAAAGCTAAAGCGGGTCAAAAAGAAAAAGA
AAGATGA
TGATGATGAACTTGACGACGAGGTGGAAGTCTGCACGCTACCGCGCCCAGGCGA
CGGGT
ACAGTGGAAGGTTCGACGCGTAAAACGTGTTTTGCGACCCGGCACCACCGTAGTC
TTTAC
GCCCCGGTGAGCGCTCCACCCGCACCTACAAGCGCGTGTATGATGAGGTGTACGGC
GACGA
GGACCTGCTTGAGCAGGCCAACGAGCGCCTCGGGGAGTTTGCCTACGGAAAGCG
GCATAA
GGACATGCTGGCGTTGCCGCTGGACGAGGGCAACCCAACACCTAGCCTAAAGCCC
GTAAC
ACTGCAGCAGGTGCTGCCCGCGCTTGCACCGTCCGAAGAAAAGCGCGGCCTAAAG
CGCGA
GTCTGGTGACTTGGCACCCACCGTGCAGCTGATGGTACCCAAGCGCCAGCGACTG
GAAGA
TGTCTTGGA AAAAATGACCGTGGAACCTGGGCTGGAGCCCGAGGTCCGCGTGCGG
CCAAT
CAAGCAGGTGGCGCCGGGACTGGGCGTGCAGACCGTGGACGTTTCAGATACCCACT
ACCAG
TAGCACCAGTATTGCCACCGCCACAGAGGGCATGGAGACACAAACGTCCCCGGTT
GCCTC
AGCGGTGGCGGATGCCGCGGTGCAGGCGGTTCGCTGCGGCCGCGTCCAAGACCTCT
ACGGA
GGTGCAAACGGACCCGTGGATGTTTCGCGTTTCAGCCCCCGGCGCCCGCGCGGT
TCGAG
GAAGTACGGCGCCGCCAGCGCGCTACTGCCGAATATGCCCTACATCCTTCCATT
GCGCC
TACCCCCGGCTATCGTGGCTACACCTACCGCCCCAGAAGACGAGCAACTACCCGA
CGCCG

AACCACCACTGGAACCCGCCGCCGCGTCGCCGTCGCCAGCCCGTGCTGGCCCCG
ATTTC
CGTGCGCAGGGTGGCTCGCGAAGGAGGCAGGACCCTGGTGCTGCCAACAGCGCG
CTACCA
CCCCAGCATCGTTTAAAAGCCGGTCTTTGTGGTTCTTGCAGATATGGCCCTCACCT
GCCG
CCTCCGTTTCCCGGTGCCGGGATTCCGAGGAAGAATGCACCGTAGGAGGGGCATG
GCCGG
CCACGGCCTGACGGGCGGCATGCGTTCGTGCGCACCACCGGCGGCGGCGCGCTCG
CACCG
TCGCATGCGCGGGCGGTATCCTGCCCTCCTTATTCCACTGATCGCCGCGGCGATTG
GCGC
CGTGCCCGGAATTGCATCCGTGGCCTTGCAGGCGCAGAGACACTGATTA AAAACA
AGTTG
CATGTGGAAAAATCAAAATAAAAAGTCTGGACTCTCACGCTCGCTTGGTCTGTGA
ACTAT
TTTGTAGAATGGAAGACATCAACTTTGCGTCTCTGGCCCCGCGACACGGCTCGCG
CCCGT
TCATGGGAAACTGGCAAGATATCGGCACCAGCAATATGAGCGGTGGCGCCTTCAG
CTGGG
GCTCGCTGTGGAGCGGCATTA AAAATTTCGGTTCCACCGTTAAGAACTATGGCAG
CAAGG
CCTGGAACAGCAGCACAGGCCAGATGCTGAGGGATAAGTTGAAAGAGCAAAATT
TCCAAC
AAAAGGTGGTAGATGGCCTGGCCTCTGGCATTAGCGGGGTGGTGGACCTGGCCAA
CCAGG
CAGTGCAAAATAAGATTAACAGTAAGCTTGATCCCCGCCCTCCCGTAGAGGAGCC
TCCAC
CGGCCGTGGAGACAGTGTCTCCAGAGGGGCGTGCGGAAAAGCGTCCGCGCCCCG
ACAGGG

AAGAAACTCTGGTGACGCAAATAGACGAGCCTCCCTCGTACGAGGAGGCACTAA
AGCAAG
GCCTGCCACCACCCGTCCCATCGCGCCCATGGCTACCGGAGTGCTGGGCCAGCA
CACAC
CCGTAACGCTGGACCTGCCTCCCCCGCCGACACCAGCAGAAACCTGTGCTGCC
AGGCC
CGACCGCCGTTGTTGTAACCCGTCCTAGCCGCGCGTCCCTGCGCCGCGCCGCCAG
CGGTC
CGCGATCGTTGCGGCCCGTAGCCAGTGGCAACTGGCAAAGCACACTGAACAGCAT
CGTGG
GTCTGGGGGTGCAATCCCTGAAGCGCCGACGATGCTTCTGAATAGCTAACGTGTC
GTATG
TGTGTCATGTATGCGTCCATGTCGCCGCCAGAGGAGCTGCTGAGCCGCCGCGCGC
CCGCT
TTCCAAGATGGCTACCCCTTCGATGATGCCGCAGTGGTCTTACATGCACATCTCGG
GCCA
GGACGCCTCGGAGTACCTGAGCCCCGGGCTGGTGCAGTTTGCCCGCGCCACCGAG
ACGTA
CTTCAGCCTGAATAACAAGTTTAGAAACCCACGGTGGCGCCTACGCACGACGTG
ACCAC
AGACCGGTCCCAGCGTTTGACGCTGCGGTTCA TCCCTGTGGACCGTGAGGATACT
GCGTA
CTCGTACAAGGCGCGGTTACCCTAGCTGTGGGTGATAACCGTGTGCTGGACATG
GCTTC
CACGTACTTTGACATCCGCGGCGTGCTGGACAGGGGCCCTACTTTTAAGCCCTACT
CTGG
CACTGCCTACAACGCCCTGGCTCCCAAGGGTGCCCCAAATCCTTGCGAATGGGAT
GAAGC
TGCTACTGCTCTTGAAATAAACCTAGAAGAAGAGGACGATGACAACGAAGACGA
AGTAGA

CGAGCAAGCTGAGCAGCAAAAACTCACGTATTTGGGCAGGCGCCTTATTCTGGT
ATAAA
TATTACAAAGGAGGGTATTCAAATAGGTGTCTGAAGGTCAAACACCTAAATATGCC
GATAA
AACATTTCAACCTGAACCTCAAATAGGAGAATCTCAGTGGTACGAAACTGAAATT
AATCA
TGCAGCTGGGAGAGTCCTTAAAAAGACTACCCCAATGAAACCATGTTACGGTTCA
TATGC
AAAACCCACAAATGAAAATGGAGGGCAAGGCATTCTTGTAAGCAACAAAATGG
AAAGCT
AGAAAGTCAAGTGGAAATGCAATTTTTCTCAACTACTGAGGCGACCGCAGGCAAT
GGTGA
TAACTTGACTCCTAAAGTGGTATTGTACAGTGAAGATGTAGATATAGAAACCCCA
GACAC
TCATATTTCTTACATGCCCACTATTAAGGAAGGTA ACTCACGAGAACTAATGGGC
CAACA
ATCTATGCCCAACAGGCCTAATTACATTGCTTTTAGGGACAATTTTATTGGTCTAA
TGTA
TTACAACAGCACGGGTAATATGGGTGTTCTGGCGGGCCAAGCATCGCAGTTGAAT
GCTGT
TGTAGATTTGCAAGACAGAAACACAGAGCTTTCATACCAGCTTTTGCTTGATTCCA
TTGG
TGATAGAACCAGGTA CTTTTCTATGTGGAATCAGGCTGTTGACAGCTATGATCCA
GATGT
TAGAATTATTGAAAATCATGGA ACTGAAGATGAACTTCCAAATTACTGCTTTCCA
CTGGG
AGGTGTGATTAATACAGAGACTCTTACCAAGGTA AAAACCTAAAACAGGTCAGGA
AAATGG
ATGGGAAAAAGATGCTACAGAATTTTCAGATAAAAATGAAATAAGAGTTGGAAA
TAATTT

TGCCATGGAAATCAATCTAAATGCCAACCTGTGGAGAAATTTCTGTACTCCAAC
ATAGC
GCTGTATTTGCCCGACAAGCTAAAGTACAGTCCTTCCAACGTAAAAATTTCTGATA
ACCC
AAACACCTACGACTACATGAACAAGCGAGTGGTGGCTCCCGGGTTAGTGGACTGC
TACAT
TAACCTTGGAGCACGCTGGTCCCTTGACTATATGGACAACGTCAACCCATTTAACC
ACCA
CCGCAATGCTGGCCTGCGCTACCGCTCAATGTTGCTGGGCAATGGTCGCTATGTGC
CCTT
CCACATCCAGGTGCCTCAGAAGTCTTTGCCATTAAAAACCTCCTTCTCCTGCCGG
GCTC
ATACACCTACGAGTGGAACCTCAGGAAGGATGTTAACATGGTTCTGCAGAGCTCC
CTAGG
AAATGACCTAAGGGTTGACGGAGCCAGCATTAAGTTTGATAGCATTTCCTTTAC
GCCAC
CTTCTCCCCATGGCCCACAACACCGCCTCCACGCTTGAGGCCATGCTTAGAAACG
ACAC
CAACGACCAGTCCTTTAACGACTATCTCTCCGCCGCAACATGCTCTACCCTATAC
CCGC
CAACGCTACCAACGTGCCCATATCCATCCCCTCCCGCAACTGGGCGGCTTTCCGCG
GCTG
GGCCTTACGCGCCTTAAGACTAAGGAAACCCCATCACTGGGCTCGGGCTACGAC
CCTTA
TTACACCTACTCTGGCTCTATACCCTACCTAGATGGAACCTTTTACCTCAACCACA
CCTT
TAAGAAGGTGGCCATTACCTTTGACTCTTCTGTCAGCTGGCCTGGCAATGACCGCC
TGCT
TACCCCAACGAGTTTGAAATTAAGCGCTCAGTTGACGGGGAGGGTTACAACGTT
GCCCA

GTGTAACATGACCAAAGACTGGTTCCTGGTACAAATGCTAGCTAACTACAACATT
GGCTA
CCAGGGCTTCTATATCCCAGAGAGCTACAAGGACCGCATGTACTCCTTCTTTAGA
AACTT
CCAGCCCATGAGCCGTCAGGTGGTGGATGATACTAAATACAAGGACTACCAACAG
GTGGG
CATCCTACACCAACACAACAACCTCTGGATTTGTTGGCTACCTTGCCCCACCATGC
GCGA
AGGACAGGCCTACCCTGCTAACTTCCCCTATCCGCTTATAGGCAAGACCGCAGTT
GACAG
CATTACCCAGAAAAAGTTTCTTTGCGATCGCACCCCTTGGCGCATCCCATTCTCCA
GTAA
CTTTATGTCCATGGGCGCACTCACAGACCTGGGCCAAAACCTTCTCTACGCCAACT
CCGC
CCACGCGCTAGACATGACTTTTGAGGTGGATCCCATGGACGAGCCCACCCTTCTTT
ATGT
TTTGTGTTGAAGTCTTTGACGTGGTCCGTGTGCACCGGCCGCACCGCGGCGTCATCG
AAAC
CGTGTACCTGCGCACGCCCTTCTCGGCCGGCAACGCCACAACATAAGCGATCGCA
GCAGG
TTTCCCCAACTGACACAAAACGTGCAACTTGAAACTCCGCCTGGTCTTTCCAGGTC
TAGA
GGGGTAACACTTTGTAAGTGGCTCCACGCTCGATCCACTGGCGAGTGTTAG
TAAC
AGCACTGTTGCTTCGTAGCGGAGCATGACGGCCGTGGGAACTCCTCCTTGGTAAC
AAGGA
CCCACGGGGCCAAAAGCCACGCCACACGGGGCCCGTCATGTGTGCAACCCCAGCA
CGGCG
ACTTTACTGCGAAACCCACTTTAAAGTGACATTGAAACTGGTACCCACACACTGG
TGACA

GGCTAAGGATGCCCTTCAGGTACCCCGAGGTAACACGCGACACTCGGGATCTGAG
AAGGG
GACTGGGGCTTCTATAAAAGCGCTCGGTTAAAAAGCTTCTATGCCTGAATANGT
GACCG
GANGTCGGCACCTTTCCTTTGCAATTAATGACCCTGTATACGCCACCATGGCTATG
ATGG
AGGTCCAGGGGGGACCCAGCCTGGGACAGACCTGCGTGCTGATCGTGATCTTTAC
AGTGC
TCCTGCAGTCTCTCTGTGTGGCTGTAACCTACGTGTACTTTACCAACGAGCTGAAG
CAGA
TGCAGGACAAGTACTCCAAAAGTGGCATTGCTTGTTTCTTAAAAGAAGATGACAG
TTATT
GGGACCCCAATGACGAAGAGAGTATGAACAGCCCCTGCTGGCAAGTCAAGTGGC
AACTCC
GTCAGCTCGTTAGAAAGATGATTTTGAGAACCTCTGAGGAAACCATTTCTACAGT
TCAAG
AAAAGCAACAAAATATTTCTCCCCTAGTGAGAGAAAGAGGTCCTCAGAGAGTAG
CAGCTC
ACATAACTGGGACCAGAGGAAGAAGCAACACATTGTCTTCTCCAAACTCCAAGAA
TGAAA
AGGCTCTGGGCCGCAAAATAAACTCCTGGGAATCATCAAGGAGTGGGCATTCATT
CCTGA
GCAACTTGCACTTGAGGAATGGTGAACCTGGTCATCCATGAAAAGGGTTTTACTA
CATCT
ATTCCAAACATACTTTCTG 认 TTTCAGGAGGAAATAAAAGAAAACACAAAGAACG
ACAAAC
AAATGGTCCAATATATTTACAAATACACAAGTTATCCTGACCCTATATTGTTGATG
AAAA
GTGCTAGA A ATAGTTGTTGGTCTAAAGATGCAGAATATGGACTCTATTCCATCTAT
CAAG

GGGGAATATTTGAGCTTAAGGAAAATGACAGAATTTTTGTTTCTGTAACAAATGA
GCACT
TAATAGACATGGACCATGAAGCCAGTTTTTTCGGGGCCTTTTTAGTTGGCTAAGTA
TACT
TCGAATGCATGCGATCGCAGAAGCAAGCAACATCAACAACAGCTGCCGCCATGG
GCTCCA
GTGAGCAGGAACTGAAAGCCATTGTCAAAGATCTTGGTTGTGGGCCATATTTTT
GGGCA
CCTATGACAAGCGCTTCCAGGCTTTGTTTCTCCACACAAGCTCGCCTGCGCCATA
GTCA
ATACGGCCGGTCGCGAGACTGGGGGCGTACACTGGATGGCCTTTCCTGGAACCC
GCACT
CAAAAACATGCTACCTCTTTGAGCCCTTTGGCTTTTCTGACCAGCGACTCAAGCAG
GTTT
ACCAGTTTGAGTACGAGTCACTCCTGCGCCGTAGCGCCATTGCTTCTTCCCCGAC
CGCT
GTATAACGCTGGAAAAGTCCACCCAAAGCGTACAGGGGCCCAACTCGGCCGCCTG
TGGAC
TATTCTGCTGCATGTTTCTCCACGCCTTTGCCAACTGGCCCCAACTCCCATGGAT
CACA
ACCCACCATGAACCTTATTACGGGGTACCCAACTCCATGCTCAACAGTCCCCA
GGTAC
AGCCCACCCTGCGTCGCAACCAGGAACAGCTCTACAGCTTCCTGGAGCGCCACTC
GCCCT
ACTTCCGCAGCCACAGTGCGCAGATTAGGAGCGCCACTTCTTTTTGTCACCTGAAA
AACA
TGTA AAAAATAATGTACTAGAGACACTTTCAATAAAGGCAAATGCTTTTATTTGT
ACACT
CTCGGGTGATTATTTACCCCCACCCTTGCCGTCTGCGCGGTTTAAAAATCAAAGGG
GTTC

TGCCGCGCATCGCTATGCGCCACTGGCAGGGACACGTTGCGATACTGGTGTTTAG
TGCTC
CACTTAAACTCAGGCACAACCATCCGCGGCAGCTCGGTGAAGTTTTCACTCCACA
GGCTG
CGCACCATCACCAACGCGTTTAGCAGGTCGGGCGCCGATATCTTGAAGTCGCAGT
TGGGG
CCTCCGCCCTGCGCGCGAGTTGCGATACACAGGGTTGCAGCACTGGAACACTA
TCAGC
GCCGGGTGGTGCACGCTGGCCAGCACGCTCTTGTCGGAGATCAGATCCGCGTCCA
GGTCC
TCCGCGTTGCTCAGGGCGAACGGAGTCAACTTTGGTAGCTGCCTTCCCAAAAAGG
GCGCG
TGCCCAGGCTTTGAGTTGCACTCGCACCGTAGTGGCATCAAAGGTGACCGTGCC
CGGTC
TGGGCGTTAGGATACAGCGCCTGCATAAAAGCCTTGATCTGCTTAAAAGCCACCT
GAGCC
TTTGCCTTCAGAGAAGAACATGCCGCAAGACTTGCCGGAAAAGTATTGGCCG
GACAG
GCCGCGTCGTGCACGCAGCACCTTGCCTCGGTGTTGGAGATCTGCACCACATTC
GGCCC
CACCGGTTCTTACGATCTTGGCCTTGCTAGACTGCTCCTTCAGCGCGCGCTGCCC
GTTT
TCGCTCGTCACATCCATTTCAATCACGTGCTCCTTATTTATCATAATGCTTCCGTGT
AGA
CACTTAAGCTCGCCTTCGATCTCAGCGCAGCGGTGCAGCCACAACGCGCAGCCCG
TGGGC
TCGTGATGCTTGTAGGTCACCTCTGCAAACGACTGCAGGTACGCCTGCAGGAATC
GCCCC
ATCATCGTCACAAAGGTCTTGTGCTGGTGAAGGTCAGCTGCAACCCGCGGTGCT
CCTCG

TTCAGCCAGGTCTTGCATACGGCCGCCAGAGCTTCCACTTGGTCAGGCAGTAGTTT
GAAG
TTCGCCTTTAGATCGTTATCCACGTGGTACTTGTCCATCAGCGCGCGCGCAGCCTC
CATG
CCCTTCTCCCACGCAGACACGATCGGCACACTCAGCGGGTTCATCACCGTAATTTT
ACTT
TCCGCTTCGCTGGGCTCTTCCTCTTCCTCTTGCCTCCGCATACCACGCGCCACTGG
GTCG
TCTTCATTCAGCCGCCGCACTGTGCGCTTACCTCCTTTGCCATGCTTGATTAGCAC
CGGT
GGGTTGCTGAAACCCACCATTGTAGCGCCACATCTTCTCTTTCTTCCTCGCTGTCC
ACG
ATTACCTCTGGTGATGGCGGGCGCTCGGGCTTGGGAGAAGGGCGCTTCTTTTTCTT
CTTG
GGCGCAATGGCCAAATCCGCCGCCGAGGTCGATGGCCGCGGGCTGGGTGTGCGC
GGCACC
AGCGCGTCTTGTGATGAGTCTTCCTCGTCCTCGGACTCGATACGCCGCCTCATCCG
CTTT
TTTGGGGGCGCCCGGGGAGGCGGCGGCGACGGGGACGGGGACGACACGTCTCTCC
ATGGTT
GGGGGACGTCGCGCCGCACCGCGTCCGCGCTCGGGGGTGGTTTCGCGCTGCTCCT
CTTCC
CGACTGGCCATTTCTTCTCCTATAGGCAGAAAAAGATCATGGAGTCAGTCGAGA
AGAAG
GACAGCCTAACCGCCCCCTCTGAGTTCGCCACCACCGCCTCCACCGATGCCGCCA
ACGCG
CCTACCACCTTCCCCGTCGAGGCACCCCCGCTTGAGGAGGAGGAAGTGATTATCG
AGCAG
GACCCAGGTTTTGTAAGCGAAGACGACGAGGACCGCTCAGTACCAACAGAGGAT
AAAAAG

CAAGACCAGGACAACGCAGAGGCAAACGAGGAACAAGTCGGGCGGGGGGACGA
AAGGCAT
GGCGACTACCTAGATGTGGGAGACGACGTGCTGTTGAAGCATCTGCAGCGCCAGT
GCGCC
ATTATCTGCGACGCGTTGCAAGAGCGCAGCGATGTGCCCTCGCCATAGCGGATG
TCAGC
CTTGCCTACGAACGCCACCTATTCTCACCGCGCGTACCCCCAAACGCCAAGAAA
ACGGC
ACATGCGAGCCCAACCCGCGCCTCAACTTCTACCCCGTATTTGCCGTGCCAGAGG
TGCTT
GCCACCTATCACATCTTTTTCCAAAACGCAAGATACCCCTATCCTGCCGTGCCAA
CCGC
AGCCGAGCGGACAAGCAGCTGGCCTTGCGGCAGGGCGCTGTCATACCTGATATCG
CCTCG
CTCAACGAAGTGCCAAAAATCTTTGAGGGTCTTGGACGCGACGAGAAGCGCGCG
GCAAAC
GCTCTGCAACAGGAAAACAGCGAAAATGAAAGTCACTCTGGAGTGTGGGTGGAA
CTCGAG
GGTGACAACGCGCGCCTAGCCGTAATAAACGCAGCATCGAGGTCACCCACTTTG
CCTAC
CCGGCACTTAACCTACCCCCAAGGTCATGAGCACAGTCATGAGTGAGCTGATCG
TGCGC
CGTGCGCAGCCCCTGGAGAGGGATGCAAATTTGCAAGAACAAACAGAGGAGGGC
CTACCC
GCAGTTGGCGACGAGCAGCTAGCGCGCTGGCTTCAAACGCGCGAGCCTGCCGACT
TGGAG
GAGCGACGCAAATAATGATGGCCGCAGTGCTCGTTACCGTGGAGCTTGAGTGCA
TGCAG
CGGTTCTTTGCTGACCCGGAGATGCAGCGCAAGCTAGAGGAAACATTGCACTACA
CCTTT

CGACAGGGCTACGTACGCCAGGCCTGCAAGATCTCCAACGTGGAGCTCTGCAACC
TGGTC
TCCTACCTTGGAATTTTGCACGAAAACCGCCTTGGGCAAACGTGCTTCATTCCAC
GCTC
AAGGGCGAGGCGCGCCGCGACTACGTCCGCGACTGCGTTTACTTATTTCTATGCT
ACACC
TGGCAGACGGCCATGGGCGTTTGGCAGCAGTGCTTGGAGGAGTGCAACCTCAAGG
AGCTG
CAGAACTGCTAAAGCAAACTTGAAGGACCTATGGACGGCCTTCAACGAGCGCT
CCGTG
GCCGCGCACCTGGCGGACATCATTTTCCCCGAACGCCTGCTTAAAACCCTGCAAC
AGGGT
CTGCCAGACTTCACCAGTCAAAGCATGTTGCAGAACTTTAGGAACTTTATCCTAG
AGCGC
TCAGGAATCTTGCCCGCCACCTGCTGTGCACTTCTAGCGACTTTGTGCCATTAA
GTAC
CGCGAATGCCCTCCGCCGCTTTGGGGCCACTGCTACCTTCTGCAGCTAGCCAACTA
CCTT
GCCTACCACTCTGACATAATGGAAGACGTGAGCGGTGACGGTCTACTGGAGTGTC
ACTGT
CGCTGCAACCTATGCACCCCGCACCGCTCCCTGGTTTGAATTCGCAGCTGCTTAA
CGAA
AGTCAAATTATCGGTACCTTTGAGCTGCAGGGTCCCTCGCCTGACGAAAAGTCCG
CGGCT
CCGGGGTTGAAACTCACTCCGGGGCTGTGGACGTTCGGCTTACCTTCGCAAATTTGT
ACCT
GAGGACTACCACGCCCACGAGATTAGGTTCTACGAAGACCAATCCCGCCCGCCAA
ATGCG
GAGCTTACCGCCTGCGTCATTACCCAGGGCCACATTCTTGGCCAATTGCAAGCCAT
CAAC

AAAGCCCGCCAAGAGTTTCTGCTACGAAAGGGACGGGGGGTTTACTTGGACCCCC
AGTCC
GGCGAGGAGCTCAACCCAATCCCCCGCCGCCGAGCCCTATCAGCAGCAGCCGC
GGGCC
CTTGCTTCCCAGGATGGCACCCAAAAAGAAGCTGCAGCTGCCGCCGCCACCCACG
GACGA
GGAGGAATACTGGGACAGTCAGGCAGAGGAGGTTTTGGACGAGGAGGAGGAGGA
CATGAT
GGAAGACTGGGAGAGCCTAGACGAGGAAGCTTCCGAGGTGGAAGAGGTGTCAGA
CGAAAC
ACCGTCACCCTCGGTGCGATTCCCCTCGCCGGCGCCCCAGAAATCGGCAACCGGT
TCCAG
CATGGCTACAACCTCCGCTCCTCAGGCGCCGCCGGCACTGCCCGTTCGCCGACCC
AACCG
TAGATGGGACACCACTGGAACCAGGGCCGGTAAGTCCAAGCAGCCGCCGCCGTT
AGCCCA
AGAGCAACAACAGCGCCAAGGCTACCGCTCATGGCGCGGGCACAAGAACGCCAT
AGTTGC
TTGCTTGCAAGACTGTGGGGGCAACATCTCCTTCGCCCGCCGCTTTCTTCTTACC
ATCA
CGGCGTGGCCTTCCCCGTAACATCCTGCATTACTACCGTCATCTCTACAGCCCAT
ACTG
CACCGGCGGCAGCGGCAGCGGCAGCAACAGCAGCGGCCACACAGAAGCAAAGGC
GACCGG
ATAGCAAGACTCTGACAAAGCCCAAGAAATCCACAGCGGCGGCAGCAGCAGGAG
GAGGAG
CGCTGCGTCTGGCGCCCAACGAACCCGTATCGACCCGCGAGCTTAGAAACAGGAT
TTTTC
CCACTCTGTATGCTATATTTCAACAGAGCAGGGGCCAAGAACAAGAGCTGAAAAT
AAAAA

ACAGGTCTCTGCGATCCCTCACCCGCAGCTGCCTGTATCACAAAAGCGAAGATCA
GCTTC
GGCGCACGCTGGAAGACGCGGAGGCTCTCTTCAGTAAATACTGCGCGCTGACTCT
TAAGG
ACTAGTTTCGCGCCCTTTCTCAAATTTAAGCGCGAAAACACTACGTCATCTCCAGCGG
CCAC
ACCCGGCGCCAGCACCTGTCGTCAGCGCCATTATGAGCAAGGAAATTCCCACGCC
CTACA
TGTGGAGTTACCAGCCACAAATGGGACTTGCGGCTGGAGCTGCCCAAGACTACTC
AACCC
GAATAAACTACATGAGCGCGGGACCCACATGATATCCCGGGTCAACGGAATCCG
CGCC
ACCGAAACCGAATTCTCTTGAACAGGCGGCTATTACCACCACACCTCGTAATAA
CCTTA
ATCCCCGTAGTTGGCCCCGCTGCCCTGGTGTACCAGGAAAGTCCCGCTCCCACCACT
GTGG
TACTTCCCAGAGACGCCAGGCCGAAGTTCAGATGACTAACTCAGGGGCGCAGCT
TGCGG
GCGGCTTTCGTCACAGGGTGC GGTCGCCCGGGCAGGGTATAACTCACCTGACAAT
CAGAG
GGCGAGGTATTCAGCTCAACGACGAGTCGGTGAGCTCCTCGCTTGGTCTCCGTCC
GGACG
GGACATTTAGATCGGCGGCGCCGGCCGTCCTTCATTCACGCCTCGTCAGGCAAT
CCTAA
CTCTGCAGACCTCGTCCTCTGAGCCGCGCTCTGGAGGCATTGGA ACTCTGCAATTT
ATTG
AGGAGTTTGTGCCATCGGTCTACTTTAACCCCTTCTCGGGACCTCCCGGCCACTAT
CCGG
ATCAATTTATTCTAACTTTGACGCGGTAAAGGACTCGGCGGACGGCTACGACTG
AATGT

TAAGTGGAGAGGCAGAGCAACTGCGCCTGAAACACCTGGTCCACTGTGCGCCGCCA
CAAGT
GCTTTGCCCGCGACTCCGGTGAGTTTTGCTACTTTGAATTGCCCGAGGATCATATC
GAGG
GCCCGGCGCACGGCGTCCGGCTTACCGCCCAGGGAGAGCTTGCCCGTAGCCTGAT
TCGGG
AGTTTACCCAGCGCCCCCTGCTAGTTGAGCGGGACAGGGGACCCTGTGTTCTCAC
TGTGA
TTTGCAACTGTCCTAACCTTGGATTACATCAAGATCTTTGTTGCCATCTCTGTGCTG
AGT
ATAATAAATACAGAAATTAATAATACTGGGGCTCCTATCGCCATCCTGTAAACG
CCACC
GTCTTCACCCGCCCAAGCAAACCAAGGCGAACCTTACCTGGTACTTTTAAACATCTC
TCCC
TCTGTGATTTACAACAGTTTCAACCCAGACGGAGTGAGTCTACGAGAGAACCTCT
CCGAG
CTCAGCTACTCCATCAGAAAAACACCACCCTCCTTACCTGCCGGGAACGTACGA
GTGCG
TCACCGGCCGCTGCACCACACCTACCGCCTGACCGTAAACCAGACTTTTTCCGGA
CAGAC
CTCAATAACTCTGTTTACCAGAACAGGAGGTGAGCTTAGAAAACCCTTAGGGTAT
TAGGC
CAAAGGCGCAGCTACTGTGGGGTTTATGAACAATTCAAGCAACTCTACGGGCTAT
TCTAA
TTCAGGTTTCTCTAGAATCGGGGTTGGGGTTATTCTCTGTCTTGTGATTCTCTTTAT
TCT
TATACTAACGCTTCTCTGCCTAAGGCTCGCCGCTGCTGTGTGCACATTTGCATTT
ATTG
TCAGCTTTTTAAACGCTGGGGTCGCCACCCAAGATGATTAGGTACATAATCCTAG
GTTTA

CTCACCCCTTGCAGTCAGCCCACGGTACCACCCAAAAGGTGGATTTTAAGGAGCCAG
CCTGT
AATGTTACATTTCGCAGCTGAAGCTAATGAGTGCACCACTCTTATAAAATGCACCA
CAGAA
CATGAAAAGCTGCTTATTCGCCACAAAACAAAATTGGCAAGTATGCTGTTTATG
CTATT
TGGCAGCCAGGTGACACTACAGAGTATAATGTTACAGTTTTCCAGGGTAAAAGTC
ATAAA
ACTTTTATGTATACTTTTCCATTTTATGAAATGTGCGACATTACCATGTACATGAG
CAAA
CAGTATAAGTTGTGGCCCCACAAAATTGTGTGGAAAACACTGGCACTTTCTGCT
GCACT
GCTATGCTAATTACAGTGCTCGCTTTGGTCTGTACCCTACTCTATATTAATACAA
AAGC
AGACGCAGCTTTATTGAGGAAAAGAAAATGCCTTAATTTACTAAGTTACAAAGCT
AATGT
CACCCTAACTGCTTTACTCGCTGCTTGCAAAAACAAATTCAAAAAGTTAGCATTAT
AATT
AGAATAGGATTTAAACCCCCGGTCATTTCCCTGCTCAATACCATTCCCCTGAACAA
TTGA
CTCTATGTGGGATATGCTCCAGCGCTACAACCTTGAAGTCAGGCTTCCCTGGATGTC
AGCA
TCTGACTTTGGCCAGCACCTGTCCCGCGGATTTGTTCCAGTCCAACCTACAGCGACC
CACC
CTAACAGAGATGACCAACACAACCAACGCGGCCCGCTACCGGACTTACATCTA
CCACA
AATACACCCCAAGTTTCTGCCTTTGTCAATAACTGGGATAACTTGGGCATGTGGTG
GTTC
TCCATAGCGCTTATGTTTGTATGCCTTATTATTATGTGGCTCATCTGCTGCCTAAAG
CGC

AAACGCGCCCGACCACCCATCTATAGTCCCATCATTGTGCTACACCCAAACAATG
ATGGA
ATCCATAGATTGGACGGACTGAAACACATGTTCTTTTCTCTTACAGTATGATTA
TGAG
ACATGATTCCTCGAGTTTTATATTACTGACCCTTGTTGCGCTTTTTTGTGCGTGCT
CCA
CATTGGCTGCGGTTTCTCACATCGAAGTAGACTGCATTCCAGCCTTCACAGTCTAT
TTGC
TTTACGGATTTGTCACCCTCACGCTCATCTGCAGCCTCATCACTGTGGTCATCGCC
TTTA
TCCAGTGCATTGACTGGGTCTGTGTGCGCTTTCATATCTCAGACACCATCCCCAG
TACA
GGGACAGGACTATAGCTGAGCTTCTTAGAATTCTTTAATTATGAAATTTACTGTGA
CTTT
TCTGCTGATTATTTGCACCCTATCTGCGTTTTGTTCCCCGACCTCCAAGCCTCAA
GACA
TATATCATGCAGATTCCTCGTATATGGAATATCCAAGTTGCTACAATGAAAA
AGCGA
TCTTTCCGAAGCCTGGTTATATGCAATCATCTCTGTTATGGTGTCTGCAGTACCA
TCTT
AGCCCTAGCTATATATCCCTACCTTGACATTGGCTGGAACGCAATAGATGCCATG
AACCA
CCCAACTTTCCCCGCGCCCGCTATGCTTCCACTGCAACAAGTTGTTGCCGGCGGCT
TTGT
CCCAGCCAATCAGCCTCGCCACCTTCTCCACCCCCACTGAAATCAGCTACTTTA
ATCT
AACAGGAGGAGATGACTGACACCCTAGATCTAGAAATGGACGGAATTATTACAG
AGCAGC
GCCTGCTAGAAAGACGCAGGGCAGCGGCCGAGCAACAGCGCATGAATCAAGAGC
TCCAAG

ACATGGTAACTTGCACCAGTGCAAAAGGGGTATCTTTTGTCTGGTAAAGCAGGC
CAAAG
TCACCTACGACAGTAATACCACCGGACACCGCCTTAGCTACAAGTTGCCAACCAA
GCGTC
AGAAATTGGTGGTCATGGTGGGAGAAAAGCCCATTACCATAACTCAGCACTCGGT
AGAAA
CCGAAGGCTGCATTCACCTCACCTTGTCAAGGACCTGAGGATCTCTGCACCCTTATT
AAGA
CCCTGTGCGGTCTCAAAGATCTTATTCCCTTTAACTAATAAAAAAAAAATAATAA
GCATC
ACTTACTTAAAATCAGTTAGCAAATTTCTGTCCAGTTTATTTCAGCAGCACCTCCTT
GCC
TCCTCCAGCTCTGGTATTGCAGCTTCCTCCTGGCTGCAAACCTTTCTCCACAATCT
AAAT
GGAATGTCAGTTTCCTCCTGTTCCGTCCATCCGCACCCACTATCTTCATGTTGTTG
CAG
ATGAAGCGCGCAAGACCGTCTGAAGATACCTTCAACCCCGTGTATCCATATGACA
CGGAA
ACCGGTCCCAACTGTGCCTTTTCTTACTCCTCCCTTTGTATCCCCCAATGGGTTT
CAA
GAGAGTCCCCCTGGGGTACTCTCTTTGCGCCTATCCGAACCTCTAGTTACCTCAA
TGGC
ATGCTTGCCTCAAAAATGGGCAACGGCCTCTCTCTGGACGAGGCCGGCAACCTTA
CCTCC
CAAAATGTAACCACTGTGAGCCCACCTCTCAAAAAACCAAGTCAAACATAAACC
TGGAA
ATATCTGCACCCCTCACAGTTACCTCAGAAGCCCTAACTGTGGCTGCCGCCGACC
TCTA
ATGGTCGCGGGCAACACACTCACCATGCAATCACAGGCCCGCTAACCGTGCACG
ACTCC

AAACTTAGCATTGCCACCCAAGGACCCCTCACAGTGTCAGAAGGAAAGCTAGCCC
TGCAA
ACATCAGGCCCCCTCACCACCACCGATAGCAGTACCCTTACTATCACTGCCTCACC
CCCT
CTAACTACTGCCACTGGTAGCTTGGGCATTGACTTGAAAGAGCCCATTTATACAC
AAAAT
GGAAAAC TAGGACTAAAGTACGGGGCTCCTTTGCATGTAACAGACGACCTAAACA
CTTTG
ACCGTAGCAACTGGTCCAGGTGTGACTATTAATAATACTTCCTTGCAAACCTAAAG
T TACT
GGAGCCTTGGGTTTTGATTCACAAGGCAATATGCAACTTAATGTAGCAGGAGGAC
TAAGG
ATTGATTCTCAAAACAGACGCCTTATACTTGATGTTAGTTATCCGTTTGATGCTCA
AAAC
CAACTAAATCTAAGACTAGGACAGGGCCCTCTTTTTATAAACTCAGCCCACAAC
TGGAT
ATTA ACTACAACAAAGGCCTTACTTGTTTACAGCTTCAAACAATTCCAAAAAGCT
TGAG
GTTAACCTAAGCACTGCCAAGGGGTTGATGTTTGACGCTACAGCCATAGCCATTA
ATGCA
GGAGATGGGCTTGAATTTGGTTCACCTAATGCACCAAACACAAATCCCCTCAAAA
CAAAA
ATTGGCCATGGCCTAGAATTTGATTCAAACAAGGCTATGGTTCCTAAACTAGGAA
CTGGC
CTTAGTTTTGACAGCACAGGTGCCATTACAGTAGGAAACAAAAATAATGATAAGC
TAACT
TTGTGGACCACACCAGCTCCATCTCCTAACTGTAGACTAAATGCAGAGAAAGATG
CTAAA
CTCACTTTGGTCTTAACAAAATGTGGCAGTCAAATACTTGCTACAGTTTCAGTTTT
GGCT

GTAAAGGCAGTTTGGCTCCAATATCTGGAACAGTTCAAAGTGCTCATCTTATTAT
AAGA
TTTGACGAAAATGGAGTGCTACTAAACAATTCCTTCCTGGACCCAGAATATTGGA
ACTTT
AGAAATGGAGATCTTACTGAAGGCACAGCCTATACAAACGCTGTTGGATTTATGC
CTAAC
CTATCAGCTTATCCAAAATCTCACGGTAAAACTGCCAAAAGTAACATTGTCAGTC
AAGTT
TACTTAAACGGAGACAAAACCTAAACCTGTAACACTAACCATTACACTAAACGGTA
CACAG
GAAACAGGAGACACAACCTCCAAGTGCATACTCTATGTCATTTTCATGGGACTGGT
CTGGC
CACAACACTACATTAATGAAATATTTGCCACATCCTCTTACACTTTTTCATACATTGC
CCAA
GAATAAAGAATCGTTTGTGTTATGTTTCAACGTGTTTATTTTTCAATTGCAGAAAA
TTTC
AAGTCATTTTTCATTCAGTAGTATAGCCCCACCACCACATAGCTTATACAGATCAC
CGTA
CCTTAATCAAACCTCACAGAACCCTAGTATTCAACCTGCCACCTCCCTCCCAACACA
CAGA
GTACACAGTCCTTTCTCCCCGGCTGGCCTTAAAAAGCATCATATCATGGGTAACA
GACAT
ATTCTTAGGTGTTATATTCCACACGGTTTCCTGTCGAGCCAAACGCTCATCAGTGA
TATT
AATAAACTCCCCGGGCAGCTCACTTAAGTTCATGTCGCTGTCCAGCTGCTGAGCC
ACAGG
CTGCTGTCCAACTTGCGGTTGCTTAACGGGCGGCGAAGGAGAAGTCCACGCCTAC
ATGGG
GGTAGAGTCATAATCGTGCATCAGGATAGGGCGGTGGTGCTGCAGCAGCGCGCG
AATAAA

CTGCTGCCGCCGCCGCTCCGTCCTGCAGGAATACAACATGGCAGTGGTCTCCTCA
GCGAT
GATTCGCACCGCCCGCAGCATAAGGCGCCTTGTCTCCGGGCACAGCAGCGCACC
CTGAT
CTCACTTAAATCAGCACAGTAACTGCAGCACAGCACCACAATATTGTTCAAAAATC
CCACA
GTGCAAGGCGCTGTATCCA A AGCTCATGGCGGGGACCACAGAACCCACGTGGCC
ATCATA
CCACAAGCGCAGGTAGATTAAGTGGCGACCCCTCATA A A CACGCTGGACATA A AC
ATTAC
CTCTTTTGGCATGTTGTAATTCACCACCTCCCGGTACCATATA A A CCTCTGATTAA
ACAT
GGCGCCATCCACCACCATCCTAAACCAGCTGGCCAAAACCTGCCCGCCGGCTATA
CACTG
CAGGGAACCGGGACTGGAACAATGACAGTGGAGAGCCCAGGACTCGTAACCATG
GATCAT
CATGCTCGTCATGATATCAATGTTGGCACAACACAGGCACACGTGCATACACTTC
CTCAG
GATTACAAGCTCCTCCCGCGTTAGAACCATATCCAGGGAACAACCCATTCCTGA
ATCAG
CGTAAATCCCACACTGCAGGGAAGACCTCGCACGTAACTCACGTTGTGCATTGTC
AAAGT
GTTACATTCGGGCAGCAGCGGATGATCCTCCAGTATGGTAGCGCGGGTTTCTGTCT
CAAA
AGGAGGTAGACGATCCCTACTGTACGGAGTGCGCCGAGACAACCGAGATCGTGTT
GGTCG
TAGTGTCATGCCAAATGGAACGCCGGACGTAGTCATATTTCCCTGAAGCAAAAACCA
GGTGC
GGGCGTGACAAACAGATCTGCGTCTCCGGTCTCGCCGCTTAGATCGCTCTGTGTA
GTAGT

TGTAGTATATCCACTCTCTCAAAGCATCCAGGCGCCCCCTGGCTTCGGGTCTATG
TAAA
TCCTTCATGCGCCGCTGCCCTGATAACATCCACCACCGCAGAATAAGCCACACC
CAGCC
AACCTACACATTCGTTCTGCGAGTCACACACGGGAGGAGCGGGAAGAGCTGGAA
GAACCA
TGTTTTTTTTTTTATTCCAAAAGATTATCCAAAACCTCAAAAATGAAGATCTATTAA
GTGA
ACGCGCTCCCCTCCGGTGGCGTGGTCAAACCTCTACAGCCAAAGAACAGATAATGG
CATT
GTAAGATGTTGCACAATGGCTTCCAAAAGGCAAACGGCCCTCACGTCCAAGTGGA
CGTAA
AGGCTAAACCCTTCAGGGTGAATCTCCTCTATAAACATTCCAGCACCTTCAACCAT
GCCC
AAATAATTCTCATCTCGCCACCTTCTCAATATATCTCTAAGCAAATCCCGAATATT
AAGT
CCGGCCATTGTAAAAATCTGCTCCAGAGCGCCCTCCACCTTCAGCCTCAAGCAGC
GAATC
ATGATTGCAAAAATTCAGGTTCCCTCACAGACCTGTATAAGATTCAAAGCGGAAC
ATTAA
CAAAAATACCGCGATCCCGTAGGTCCCTTCGCAGGGCCAGCTGAACATAATCGTG
CAGGT
CTGCACGGACCAGCGCGGCCACTTCCCCGCCAGGAACCTTGACAAAAGAACCCAC
ACTGA
TTATGACACGCATACTCGGAGCTATGCTAACAGCGTAGCCCCGATGTAAGCTTT
GTTGC
ATGGGCGGCGATATAAAAATGCAAGGTGCTGCTCAAAAATCAGGCAAAGCCTCG
CGCAA
AAAGAAAGCACATCGTAGTCATGCTCATGCAGATAAAGGCAGGTAAGCTCCGGA
ACCACC

ACAGAAAAGACACCATTTTTCTCTCAAACATGTCTGCGGGTTTCTGCATAAACA
CAAAA
TAAAATAACAAAAAACATTTAAACATTAGAAGCCTGTCTTACAACAGGAAAAA
CAACCC
TTATAAGCATAAGACGGACTACGGCCATGCCGGCGTGACCGTAAAAAACTGGTC
ACCGT
GATTA AAAAGCACCACCGACAGCTCCTCGGTCATGTCCGGAGTCATAATGTAAGA
CTCGG
TAAACACATCAGGTTGATTCATCGGTCAGTGCTAAAAAGCGACCGAAATAGCCCG
GGGA
ATACATACCCGCAGGCGTAGAGACAACATTACAGCCCCATAGGAGGTATAACA
AAATTA
ATAGGAGAGAAAAACACATAAACACCTGAAAAACCTCCTGCCTAGGCAAATA
GCACCC
TCCCGCTCCAGAACAACATACAGCGCTTCCACAGCGGCAGCCATAACAGTCAGCC
TTACC
AGTAAAAAAGAAAACCTATTA AAAAACACCACTCGACACGGCACCAGCTCAAT
CAGTCA
CAGTGTA AAAAAGGGCCAAGTGCAGAGCGAGTATATATAGGACTAAAAAATGAC
GTAACG
GTAAAGTCCACAAAAACACCCAGAAAACCGCACGCGAACCTACGCCAGAAA
CGAAAG
CCAAAAACCCACA ACTTCTCAAATCGTCACTTCCGTTTTCCACGTTACGTCAC
TTCC
CATTTTAATTAAGAAA ACTACAATTCCCAACACATAACAAGTTACTCCGCCCTAAA
ACCTA
CGTACCCGCCCCGTTCCACGCCCCGCGCCACGTCACAAACTCCACCCCCTCATT
ATCA
TATTGGCTTCAATCCAAAATAAGGTATATTATTGATGATGATTACCCTGTTAT
SEQ ID NO: 51 是 OV1160 序列

AGGGTAATCATCATCAATAATATACCTTATTTTGGATTGAAGCCAATATGATAATG
AGGG
GGTGGAGTTTGTGACGTGGCGCGGGGCGTGGGAACGGGGCGGGTGACGTAGTAG
TGTGGC
GGAAGTGTGATGTTGCAAGTGTGGCGGAACACATGTAAGCGACGGATGTGGCAA
AAGTGA
CGTTTTTGGTGTGCGCCGGTGTACACAGGAAGTGACAATTTTCGCGCGGTTTTAGG
CGGA
TGTTGTAGTAAATTTGGGCGTAACCGAGTAAGATTTGGCCATTTTCGCGGGAAAA
CTGAA
TAAGAGGAAGTGAAATCTGAATAATTTTGTGTTACTCATAGCGCGTAATATTTGTC
TAGG
GCCGGGATCTCTGCAGGAATTTGATATCAAGCTTATCGATACCGTCGAAACTTGT
TATT
GCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTCACAAATAAAG
CATTT
TTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTC
TGG
ATCCGCTAGCGGCGCGCCGTTTCATCCGGACAAAGCCTGCGCGCGCCCCGCCCG
CCATT
GGCCGTACCGCCCCGCGCCGCCGCCCATCTCGCCCCTCGCCGCCGGGTCCGGCG
CGTTA
AAGCCAATAGGAACCGCCCGCTTGTTCCTCGTCACGGCCGGGGCAGCCAATTGTG
GCGGC
GCTCGGCGGCTCGTGGCTCTTTCGCGGCAAAAAGGATTTGGCGCGTAAAAGTGGC
CGGGA
CTTTGCAGGCAGCGGCGGCCGGGGGCGGAGCGGGATCGAGCCCTCGATGATATC
AGATCA
AACGATATCACCGGTCGACTGAAAATGAGACATATTATCTGCCACGGAGGTGTTA
TTACC

GAAGAAATGGCCGCCAGTCTTTTGGACCAGCTGATCGAAGAGGTA CTGGCTGATA
ATCTT
CCACCTCCTAGCCATTTTGAACCACCTACCCTTCACGAACTGTATGATTTAGACGT
GACG
GCCCCGAAGATCCCAACGAGGAGGCGGTTTCGCAGATTTTTCCCGACTCTGTAA
TGTTG
GCGGTGCAGGAAGGGATTGACTTACTCACTTTTCCGCCGGCGCCCGGTTCTCCGG
AGCCG
CCTCACCTTTCCCGGCAGCCCGAGCAGCCGGAGCAGAGAGCCTTGGGTCCGGTTT
CTATG
CCAAACCTTGTACCGGAGGTGATCGATCTTACCTGCCACGAGGCTGGCTTCCACC
CAGT
GACGACGAGGATGAAGAGGGTGAGGAGTTTGTGTTAGATTATGTGGAGCACCCC
GGGCAC
GGTTGCAGGTCTTGTCATTATCACCGGAGGAATACGGGGGACCCAGATATTATGT
GTTCCG
CTTTGCTATATGAGGACCTGTGGCATGTTTGTCTACAGTAAGTGAAAATTATGGGC
AGTG
GGTGATAGAGTGGTGGGTTTGGTGTGGTAATTTTTTTTTTAATTTTACAGTTTGT
GGT
TTAAAGAATTTTGTATTGTGATTTTTTTAAAAGGTCTGTGTCTGAACCTGAGCCT
GAGC
CCGAGCCAGAACCGGAGCCTGCAAGACCTACCCGCCGTCCTAAAATGGCGCCTGC
TATCC
TGAGACGCCCACATCACCTGTGTCTAGAGAATGCAATAGTAGTACGGATAGCTG
TGACT
CCGGTCCTTCTAACACACCTCCTGAGATACACCCGGTGGTCCCGCTGTGCCCCATT
AAAC
CAGTTGCCGTGAGAGTTGGTGGGCGTCGCCAGGCTGTGGAAATGTATCGAGGACTT
GCTTA

ACGAGCCTGGGCAACCTTTGGACTTGAGCTGTAAACGCCCCAGGCCATAAGGTGT
AAACC
TGTGATTGCGTGTGTGGTTAACGCCTTTGTTTGCTGAATGAGTTGATGTAAGTTTA
ATAA
AGGGTGAGATAATGTTTAACTTGCATGGCGTGTAAATGGGGCGGGGCTTAAAGG
GTATA
TAATGCGCCGTGGGCTAATCTTGGTTACATCTGACCTCATGGAGGCTTGGGAGTGT
TTGG
AAGATTTTTCTGCTGTGCGTAACTTGTGGAACAGAGCTCTAACAGTACCTCTTGG
TTTT
GGAGGTTTCTGTGGGGCTCATCCCAGGCAAAGTTAGTCTGCAGAATTAAGGAGGA
TTACA
AGTGGGAATTTGAAGAGCTTTTGAAATCCTGTGGTGAGCTGTTTGATTCTTTGAAT
CTGG
GTCACCAGGCGCTTTTCCAAGAGAAGGTCATCAAGACTTTGGATTTTTCCACACCG
GGGC
GCGCTGCGGCTGCTGTTGCTTTTTTTGAGTTTTATAAAGGATAAATGGAGCGAAGA
AACCC
ATCTGAGCGGGGGGTACCTGCTGGATTTTCTGGCCATGCATCTGTGGAGAGCGGT
TGTGA
GACACAAGAATCGCCTGCTACTGTTGTCTTCCGTCCGCCCGGCGATAATACCGAC
GGAGG
AGCAGCAGCAGCAGCAGGAGGAAGCCAGGCGGGCGGCAGGAGCAGAGCCCA
TGGAACC
CGAGAGCCGGCCTGGACCCTCGGGAATGAATGTTGTACAGGTGGCTGAACTGTAT
CCAGA
ACTGAGACGCATTTTGACAATTACAGAGGATGGGCAGGGGCTAAAGGGGGTAAA
GAGGGA
GCGGGGGGCTTGTGAGGCTACAGAGGAGGCTAGGAATCTAGCTTTTAGCTTAATG
ACCAG

ACACCGTCCTGAGTGTATTACTTTTCAACAGATCAAGGATAATTGCGCTAATGAG
CTTGA
TCTGCTGGCGCAGAAGTATTCCATAGAGCAGCTGACCACTTACTGGCTGCAGCCA
GGGGA
TGATTTTGAGGAGGCTATTAGGGTATATGCAAAGGTGGCACTTAGGCCAGATTGC
AAGTA
CAAGATCAGCAAACCTTGTAATATCAGGAATTGTTGCTACATTTCTGGGAACGGG
GCCGA
GGTGGAGATAGATACGGAGGATAGGGTGGCCTTTAGATGTAGCATGATAAATATG
TGGCC
GGGGGTGCTTGGCATGGACGGGGTGGTTATTATGAATGTAAGGTTTACTGGCCCC
AATTT
TAGCGGTACGGTTTTCTGGCCAATACCAACCTTATCCTACACGGTGTAAGCTTCT
ATGG
GTTTAACAATACCTGTGTGGAAGCCTGGACCGATGTAAGGGTTCGGGGCTGTGCC
TTTTA
CTGCTGCTGGAAGGGGGTGGTGTGTCGCCCCAAAAGCAGGGCTCAATTAAGAAA
TGCCT
CTTTGAAAGGTGTACCTTGGGTATCCTGTCTGAGGGTAACTCCAGGGTGCGCCAC
AATGT
GGCCTCCGACTGTGGTTGCTTCATGCTAGTGAAAAGCGTGGCTGTGATTAAGCAT
AACAT
GGTATGTGGCAACTGCGAGGACAGGGCCTCTCAGATGCTGACCTGCTCGGACGGC
AACTG
TCACCTGCTGAAGACCATTACGTAGCCAGCCACTCTCGCAAGGCCTGGCCAGTG
TTTGA
GCATAACATACTGACCCGCTGTTCCCTTGCATTTGGGTAACAGGAGGGGGGTGTTT
CTACC
TTACCAATGCAATTTGAGTCACACTAAGATATTGCTTGAGCCCGAGAGCATGTCC
AAGGT

GAACCTGAACGGGGTGTTTGACATGACCATGAAGATCTGGAAGGTGCTGAGGTAC
GATGA
GACCCGCACCAGGTGCAGACCCTGCGAGTGTGGCGGTAAACATATTAGGAACCA
GCCTGT
GATGCTGGATGTGACCGAGGAGCTGAGGCCCGATCACTTGGTGCTGGCCTGCACC
CGCGC
TGAGTTTGGCTCTAGCGATGAAGATACAGATTGAGGTAAGTAAATGTGTGGGCGT
GGCTT
AAGGGTGGGAAAGAATATATAAGGTGGGGGTCTTATGTAGTTTTGTATCTGTTTT
GCAGC
AGCCGCCGCCCATGAGCACCAACTCGTTTGATGGAAGCATTGTGAGCTCATAT
TTGAC
AACGCGCATGCCCCATGGGCCGGGGTGCGTCAGAATGTGATGGGCTCCAGCATT
GATGG
TCGCCCCGTCTGCCCGCAAACCTACTACCTTGACCTACGAGACCGTGTCTGGAA
CGCC
GTTGGAGACTGCAGCCTCCGCCGCCGCTTCAGCCGCTGCAGCCACCGCCCGCGGG
ATTGT
GACTGACTTTGCTTTCCTGAGCCCGCTTGCAAGCAGTGCAGCTTCCCGTTCATCCG
CCCG
CGATGACAAGTTGACGGCTCTTTTGGCACAATTGGATTCTTTGACCCGGGAACCTTA
ATGT
CGTTTCTCAGCAGCTGTTGGATCTGCGCCAGCAGGTTTCTGCCCTGAAGGCTTCCT
CCCC
TCCCAATGCGGTTTAAAACATAAATAAAAAACCAGACTCTGTTTGGATTTGGATC
AAGCA
AGTGTCTTGCTGTCTTTATTTAGGGGTTTTGCGCGCGCGGTAGGCCCGGGACCAGC
GGTC
TCGGTCGTTGAGGGTCCCTGTGTATTTTTTCCAGGACGTGGTAAAGGTGACTCTGGA
TGTT

CAGATACATGGGCATAAGCCCGTCTCTGGGGTGGAGGTAGCACCCTGCAGAGCT
TCATG
CTGCGGGGTGGTGTGTAGATGATCCAGTCGTAGCAGGAGCGCTGGGCGTGGTGC
CTAAA
AATGTCTTTCAGTAGCAAGCTGATTGCCAGGGGCAGGCCCTTGGTGTAAAGTGTTT
ACAAA
GCGGTTAAGCTGGGATGGGTGCATACGTGGGGATATGAGATGCATCTTGGACTGT
ATTTT
TAGGTTGGCTATGTTCCCAGCCATATCCCTCCGGGGATTCATGTTGTGCAGAACCA
CCAG
CACAGTGTATCCGGTGCACCTGGGAAATTTGTCATGTAGCTTAGAAGGAAATGCG
TGGAA
GAACTTGGAGACGCCCTTGTGACCTCCAAGATTTTCCATGCATTCGTCCATAATGA
TGGC
AATGGGCCACGGGCGGCGGCCTGGGCGAAGATATTTCTGGGATCACTAACGTCA
TAGTT
GTGTTCCAGGATGAGATCGTCATAGGCCATTTTTACAAAGCGCGGGCGGAGGGTG
CCAGA
CTGCGGTATAATGGTTCATCCGGCCCAGGGGCGTAGTTACCCTCACAGATTGTC
ATTTC
CCACGCTTTGAGTTCAGATGGGGGGATCATGTCTACCTGCGGGGCGATGAAGAAA
ACGGT
TTCCGGGGTAGGGGAGATCAGCTGGGAAGAAAGCAGGTTCTGAGCAGCTGCGA
CTTACC
GCAGCCGGTGGGCCCGTAAATCACACCTATTACCGGGTGCAACTGGTAGTTAAGA
GAGCT
GCAGCTGCCGTCATCCCTGAGCAGGGGGGCCACTTCGTTAAGCATGTCCCTGACT
CGCAT
GTTTTCCCTGACCAAATCCGCCAGAAGGCGCTCGCCGCCAGCGATAGCAGTTCT
TGCAA

GGAAGCAAAGTTTTTCAACGGTTTGAGACCGTCCGCCGTAGGCATGCTTTTGAGC
GTTTG
ACCAAGCAGTTCCAGGCGGTCCCACAGCTCGGTACCTGCTCTACGGCATCTCGA
TCCAG
CATATCTCCTCGTTTCGCGGGTTGGGGCGGCTTTCGCTGTACGGCAGTAGTCGGTG
CTCG
TCCAGACGGGCCAGGGTCATGTCTTCCACGGGCGCAGGGTCCTCGTCAGCGTAG
TCTGG
GTCACGGTGAAGGGGTGCGCTCCGGGCTGCGCGCTGGCCAGGGTGCCTTGAGGC
TGGTC
CTGCTGGTGCTGAAGCGCTGCCGGTCTTCGCCCTGCGCGTCGGCCAGGTAGCATT
GACC
ATGGTGTACATAGTCCAGCCCCTCCGCGGCGTGGCCCTTGGCGCGCAGCTTGCCCTT
GGAG
GAGGCGCCGCACGAGGGGCAGTGCAGACTTTTGAGGGCGTAGAGCTTGGGCGCG
AGAAAT
ACCGATTCCGGGGAGTAGGCATCCGCGCCGCAGGCCCCGCAGACGGTCTCGCATT
CCACG
AGCCAGGTGAGCTCTGGCCGTTCCGGGTCAAAAACCAGGTTTCCCCATGCTTTTT
GATG
CGTTTCTTACCTCTGGTTTCCATGAGCCGGTGTCCACGCTCGGTGACGAAAAGGCT
GTCC
GTGTCCCCGTATACAGACTTGAGAGGCCTGTCTCGAGCGGTGTTCCGCGGTCTC
CTCG
TATAGAAACTCGGACCACTCTGAGACAAAGGCTCGCGTCCAGGCCAGCACGAAG
GAGGCT
AAGTGGGAGGGGTAGCGGTCGTTGTCCACTAGGGGGTCCACTCGCTCCAGGGTGT
GAAGA
CACATGTCGCCCTCTTCGGCATCAAGGAAGGTGATTGGTTTGTAGGTGTAGGCCA
CGTGA

CCGGGTGTTCTGAAGGGGGGCTATAAAAGGGGGTGGGGGCGCGTTTCGTCCTCAC
TCTCT
TCCGCATCGCTGTCTGCGAGGGCCAGCTGTTGGGGTGAGTACTCCCTCTGAAAAG
CGGGC
ATGACTTCTGCGCTAAGATTGTCAGTTTCCAAAACGAGGAGGATTTGATATTCA
CCTGG
CCCGCGGTGATGCCTTTGAGGGTGGCCGCATCCATCTGGTCAGAAAAGACAATCT
TTTTG
TTGTCAAGCTTGGTGGCAAACGACCCGTAGAGGGCGTTGGACAGCAACTTGGCGA
TGGAG
CGCAGGGTTTGGTTTTTGTGCGGATCGGCGCGCTCCTTGGCCGCGATGTTTAGCTG
CACG
TATTCGCGCGCAACGCACCGCCATTCGGGAAAGACGGTGGTGCCTCGTCGGGCA
CCAGG
TGCACGCGCCAACCGCGGTTGTGCAGGGTGACAAGGTCAACGCTGGTGGCTACCT
CTCCG
CGTAGGCGCTCGTTGGTCCAGCAGAGGCGGCCGCCCTTGC GCGAGCAGAATGGCG
GTAGG
GGGTCTAGCTGCGTCTCGTCCGGGGGTCTGCGTCCACGGTAAAGACCCCGGGCA
GCAGG
CGCGGTCGAAGTAGTCTATCTTGCAATCCTTGCAAGTCTAGCGCCTGCTGCCATGC
GCGG
GCGGCAAGCGCGCGCTCGTATGGGTTGAGTGGGGGACCCCATGGCATGGGGTGG
GTGAGC
GCGGAGGCGTACATGCCGCAAATGTCGTAAACGTAGAGGGGCTCTCTGAGTATTC
CAAGA
TATGTAGGGTAGCATCTTCCACCGCGGATGCTGGCGCGCACGTAATCGTATAGTT
CGTGC
GAGGGAGCGAGGAGGTCGGGACCGAGGTTGCTACGGGCGGGCTGCTCTGCTCGG
AAGACT

ATCTGCCTGAAGATGGCATGTGAGTTGGATGATATGGTTGGACGCTGGAAGACGT
TGAAG
CTGGCGTCTGTGAGACCTACCGCGTCACGCACGAAGGAGGCGTAGGAGTCGCGCA
GCTTG
TTGACCAGCTCGGCGGTGACCTGCACGTCTAGGGCGCAGTAGTCCAGGGTTTCCT
TGATG
ATGTCATACTTATCCTGTCCCTTTTTTTTCCACAGCTCGCGGTTGAGGACAAACTCT
TCG
CGGTCTTTCCAGTACTCTTGGATCGGAAACCCGTCGGCCTCCGAACGGTAAGAGC
CTAGC
ATGTAGAACTGGTTGACGGCCTGGTAGGCGCAGCATCCCTTTTCTACGGGTAGCG
CGTAT
GCCTGCGCGGCCTTCCGGAGCGAGGTGTGGGTGAGCGCAAAGGTGTCCCTGACCA
TGACT
TTGAGGTACTGGTATTTGAAGTCAGTGTGTCGTCGCATCCGCCCTGCTCCCAGAGCAA
AAAG
TCCGTGCGCTTTTTGGAACGCGGATTTGGCAGGGCGAAGGTGACATCGTTGAAGA
GTATC
TTTCCCGCGCGAGGCATAAAGTTGCGTGTGATGCGGAAGGGTCCCGGCACCTCGG
AACGG
TTGTTAATTACCTGGGCGGCGAGCACGATCTCGTCAAAGCCGTTGATGTTGTGGCC
CACA
ATGTAAAGTTCCAAGAAGCGCGGGATGCCCTTGATGGAAGGCAATTTTTTAAGTT
CCTCG
TAGGTGAGCTCTTCAGGGGAGCTGAGCCCGTGCTCTGAAAGGGCCCAGTCTGCAA
GATGA
GGGTTGGAAGCGACGAATGAGCTCCACAGGTCACGGGCCATTAGCATTGTCAGGT
GGTCG
CGAAAGGTCCTAAACTGGCGACCTATGGCCATTTTTTCTGGGGTGATGCAGTAGA
AGGTA

AGCGGGTCTTGTCCAGCGGTCCCATCCAAGGTTTCGCGGCTAGGTCTCGCGCGG
CAGTC
ACTAGAGGCTCATCTCCGCCGAACCTCATGACCAGCATGAAGGGCACGAGCTGCT
TCCCA
AAGGCCCCCATCCAAGTATAGGTCTCTACATCGTAGGTGACAAAGAGACGCTCGG
TGCGA
GGATGCGAGCCGATCGGGAAGAAGTGGATCTCCCGCCACCAATTGGAGGAGTGG
CTATTG
ATGTGGTGAAAGTAGAAGTCCCTGCGACGGGCCGAACACTCGTGCTGGCTTTTGT
AAAAA
CGTGCGCAGTACTGGCAGCGGTGCACGGGCTGTACATCCTGCACGAGGTTGACCT
GACGA
CCGCGCACAAGGAAGCAGAGTGGGAATTTGAGCCCCTCGCCTGGCGGGTTTGGCT
GGTGG
TCTTCTACTTCGGCTGCTTGTCTTGACCGTCTGGCTGCTCGAGGGGAGTTACGGT
GGAT
CGGACCACCACGCCGCGGAGCCCAAAGTCCAGATGTCCGCGCGCGGGCGGTCGG
AGCTTG
ATGACAACATCGCGCAGATGGGAGCTGTCCATGGTCTGGAGCTCCCGCGGCGTCA
GGTCA
GGCGGGAGCTCCTGCAGGTTTACCTCGCATAGACGGGTCAGGGCGCGGGGCTAGAT
CCAGG
TGATACCTAATTTCCAGGGGCTGGTTGGTGGCGGCGTCGATGGCTTGCAAGAGGC
CGCAT
CCCCGCGGCGGACTACGGTACCGCGCGGCGGGCGGTGGGCCGCGGGGGTGTCTT
TGGAT
GATGCATCTAAAAGCGGTGACGCGGGCGAGCCCCGGAGGTAGGGGGGGCTCCG
GACCCG
CCGGGAGAGGGGGCAGGGGCACGTCCGGCGCCGCGCGGGCAGGAGCTGGTGCT
GCGCGC

GTAGGTTGCTGGCGAACGCGACGACGCGGGCGGTTGATCTCCTGAATCTGGCGCCT
CTGCG
TGAAGACGACGGGCCCCGGTGAGCTTGAGCCTGAAAGAGAGTTCGACAGAATCAA
TTTCGG
TGTCGTTGACGGCGGCCTGGCGCAAATCTCCTGCACGTCTCCTGAGTTGTCTTGA
TAGG
CGATCTCGGCCATGAACTGCTCGATCTCTTCTCCTGGAGATCTCCGCGTCCGGCT
CGCT
CCACGGTGGCGGGCAGGTCGTTGAAATGCGGGCCATGAGCTGCGAGAAGGCGT
TGAGGC
CTCCCTCGTTCCAGACGCGGCTGTAGACCACGCCCCCTTCGGCATCGCGGGCGCG
CATGA
CCACCTGCGCGAGATTGAGCTCCACGTGCCGGGCGAAGACGGCGTAGTTTCGCAG
GCGCT
GAAAGAGGTAGTTGAGGGTGGTGGCGGTGTGTTCTGCCACGAAGAAGTACATAA
CCCAGC
GTCGCAACGTGGATTTCGTTGATATCCCCAAGGCTCAAGGCGCTCCATGGCCTC
GTAGA
AGTCCACGGCGAAGTTGAAAACTGGGAGTTGCGCGCCGACACGGTTAACTCCTC
CTCCA
GAAGACGGATGAGCTCGGCGACAGTGTGCGGCACCTCGCGCTCAAAGGCTACAG
GGGCCT
CTTCTTCTTCAATCTCCTCTTCCATAAGGGCCTCCCCTTCTTCTTCTTGGCG
GCG
GTGGGGGAGGGGGACACGGCGGGCAGACGGCGCACCGGGAGGCGGTGACAA
AAGCGCT
CGATCATCTCCCCGCGGGCAGGGCGCATGGTCTCGGTGACGGCGCGGCCGTTCTC
GCGGG
GGCGCAGTTGGAAGACGCCGCCCGTCATGTCCCGTTATGGGTTGGCGGGGGGCT
GCCAT

GCGGCAGGGATACGGCGCTAACGATGCATCTCAACAATTGTTGTGTAGGTA CTCC
GCCGC
CGAGGGACCTGAGCGAGTCCGCATCGACCGGATCGGAAAACCTCTCGAGAAAGG
CGTCTA
ACCAGTCACAGTCGCAAGGTAGGCTGAGCACCGTGGCGGGCGGCAGCGGGCGGC
GGTCGG
GGTTGTTTCTGGCGGAGGTGCTGCTGATGATGTAATTAAGTAGGCGGTCTTGAG
ACGGC
GGATGGTCGACAGAAGCACCATGTCCTTGGGTCCGGCCTGCTGAATGCGCAGGCG
GTCGG
CCATGCCCCAGGCTTCGTTTTGACATCGGCGCAGGTCTTTGTAGTAGTCTTG CATG
AGCC
TTTCTACCGGCACTTCTTCTTCCTTCTTCTGCTGCATCTCTTG CATCTATCGC
TG
CGGCGGGCGGCGGAGTTTGGCCGTAGGTGGCGCCCTCTTCTCCCATGCGTGTGAC
CCCGA
AGCCCCTCATCGGCTGAAGCAGGGCTAGGTCCGGCGACAACGCGCTCGGCTAATAT
GGCCT
GCTGCACCTGCGTGAGGGTAGACTGGAAGTCATCCATGTCCACAAAGCGGTGGTA
TGCGC
CCGTGTTGATGGTGTAAGTGCAGTTGGCCATAACGGACCAGTTAACGGTCTGGTG
ACCCG
GCTGCGAGAGCTCGGTGTACCTGAGACGCGAGTAAGCCCTCGAGTCAAATACGTA
GTCGT
TGCAAGTCCGCACCAGGTACTGGTATCCCACCAAAAAGTGCGGGCGGCGGCTGGCG
GTAGA
GGGGCCAGCGTAGGGTGGCCGGGGCTCCGGGGGCGAGATCTTCCAACATAAGGC
GATGAT
ATCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCG
CGGAA

AGTCGCGGACGCGGTTCCAGATGTTGCGCAGCGGCAAAAAGTGCTCCATGGTCGG
GACGC
TCTGGCCGGTCAGGCGCGCGCAATCGTTGACGCTCTAGACCGTGCAAAAGGAGAG
CCTGT
AAGCGGGCACTCTTCCGTGGTCTGGTGGATAAATTCGCAAGGGTATCATGGCGGA
CGACC
GGGGTTCGAGCCCCGTATCCGGCCGTCCGCCGTGATCCATGCGGTTACCGCCCCG
GTGTC
GAACCCAGGTGTGCGACGTCAGACAACGGGGGAGTGCTCCTTTTGGCTTCCTTCC
AGGCG
CGGCGGCTGCTGCGCTAGCTTTTTTGGCCACTGGCCGCGCGCAGCGTAAGCGGTT
AGGCT
GGAAAGCGAAAGCATTAAAGTGGCTCGCTCCCTGTAGCCGGAGGGTTATTTTCCAA
GGTT
GAGTCGCGGGACCCCCGGTTCGAGTCTCGGACCGGCCGGACTGCGGCGAACGGG
GGTTTG
CCTCCCCGTCATGCAAGACCCCGCTTGCAAATTCCTCCGAAACAGGGACGAGCC
CCTTT
TTTGCTTTTCCAGATGCATCCGGTGCTGCGGCAGATGCGCCCCCTCCTCAGCAG
CGGC
AAGAGCAAGAGCAGCGGCAGACATGCAGGGCACCCCTCCCCTCCTCCTACCGCGTC
AGGAG
GGGCGACATCCGCGGTTGACGCGGCAGCAGATGGTGATTACGAACCCCCGCGGC
GCCGGG
CCCGGCACTACCTGGACTTGGAGGAGGGCGAGGGCCTGGCGCGGCTAGGAGCGC
CCTCTC
CTGAGCGGTACCCAAGGGTGCAGCTGAAGCGTGATACGCGTGAGGCGTACGTGCC
GCGGC
AGAACCTGTTTCGCGACCGCGAGGGAGAGGGAGCCCGAGGAGATGCGGGATCGAA
AGTTCC

ACGCAGGGCGCGAGCTGCGGCATGGCCTGAATCGCGAGCGGTTGCTGCGCGAGG
AGGACT
TTGAGCCCGACGCGCGAACCGGGATTAGTCCCGCGCGCGCACACGTGGCGGCCGC
CGACC
TGGTAACCGCATAACGAGCAGACGGTGAACCAGGAGATTAACTTTCAAAAAAGCTT
TAACA
ACCACGTGCGTACGCTTGTGGCGCGGAGGAGGTGGCTATAGGACTGATGCATCT
GTGGG
ACTTTGTAAGCGCGCTGGAGCAAACCCAAATAGCAAGCCGCTCATGGCGCAGCT
GTTCC
TTATAGTGCAGCACAGCAGGGACAACGAGGCATTCAGGGATGCGCTGCTAAACAT
AGTAG
AGCCCGAGGGCCGCTGGCTGCTCGATTTGATAAACATCCTGCAGAGCATAGTGGT
GCAGG
AGCGCAGCTTGAGCCTGGCTGACAAGGTGGCCGCCATCAACTATTCCATGCTTAG
CCTGG
GCAAGTTTTACGCCCGCAAGATATAACCATACCCCTTACGTTCCCATAGACAAGGA
GGTAA
AGATCGAGGGGTTCTACATGCGCATGGCGCTGAAGGTGCTTACCTTGAGCGACGA
CCTGG
GCGTTTATCGCAACGAGCGCATCCACAAGGCCGTGAGCGTGAGCCGGCGGCGCG
AGCTCA
GCGACCGCGAGCTGATGCACAGCCTGCAAAGGGCCCTGGCTGGCACGGGCAGCG
GCGATA
GAGAGGCCGAGTCCTACTTTGACGCGGGCGCTGACCTGCGCTGGGCCCAAGCCG
ACGCG
CCCTGGAGGCAGCTGGGGCCGGACCTGGGCTGGCGGTGGCACCCCGCGCGCGCTG
GCAACG
TCGGCGGCGTGGAGGAATATGACGAGGACGATGAGTACGAGCCAGAGGACGGCG
AGTACT

AAGCGGTGATGTTTCTGATCAGATGATGCAAGACGCAACGGACCCGGCGGTGCGG
GCGGC
GCTGCAGAGCCAGCCGTCCGGCCTTAACTCCACGGACGACTGGCGCCAGGTCATG
GACCG
CATCATGTGCTGACTGCGCGCAATCCTGACGCGTTCCGGCAGCAGCCGCAGGCC
AACCG
GCTCTCCGCAATTCTGGAAGCGGTGGTCCCGGCGCGCGCAAACCCACGCACGAG
AAGGT
GCTGGCGATCGTAAACGCGCTGGCCGAAAACAGGGCCATCCGGCCCGACGAGGC
CGGCCT
GGTCTACGACGCGCTGCTTCAGCGCGTGGCTCGTTACAACAGCGGCAACGTGCAG
ACCAA
CCTGGACCGGCTGGTGGGGGATGTGCGCGAGGCCGTGGCGCAGCGTGAGCGCGC
GCAGCA
GCAGGGCAACCTGGGCTCCATGGTTGCACTAAACGCCTTCCTGAGTACACAGCCC
GCCAA
CGTGCCGCGGGGACAGGAGGACTACACCAACTTTGTGAGCGCACTGCGGCTAATG
GTGAC
TGAGACACCGCAAAGTGAGGTGTACCAGTCTGGGCCAGACTATTTTTTCCAGACC
AGTAG
ACAAGGCCTGCAGACCGTAAACCTGAGCCAGGCTTTCAAAAATTGCAGGGGCTG
TGGGG
GGTGCGGGCTCCACAGGCGACCGCGCGACCGTGTCTAGCTTGCTGACGCCAAC
TCGCG
CCTGTTGCTGCTGCTAATAGCGCCCTTCACGGACAGTGGCAGCGTGTCCCGGGAC
ACATA
CCTAGGTCACTTGCTGACACTGTACCGCGAGGCCATAGGTCAGGCGCATGTGGAC
GAGCA
TACTTTCCAGGAGATTACAAGTGTGACCCGCGCGCTGGGGCAGGAGGACACGGGC
AGCCT

GGAGGCAACCCTAAACTACCTGCTGACCAACCGGCGGCAGAAGATCCCCTCGTTG
CACAG
TTTAAACAGCGAGGAGGAGCGCATTTTGCCTACGTGCAGCAGAGCGTGAGCCTT
AACCT
GATGCGCGACGGGGTAACGCCAGCGTGGCGCTGGACATGACCGCGCGCAACAT
GGAACC
GGGCATGTATGCCTCAAACCGGCCGTTTATCAACCGCCTAATGGACTACTTGCATC
GCGC
GGCCGCCGTGAACCCCGAGTATTTACCAATGCCATCTTGAACCCGCACTGGCTA
CCGCC
CCCTGGTTTCTACACCGGGGGATTTCGAGGTGCCCGAGGGTAACGATGGATTCCTC
TGGGA
CGACATAGACGACAGCGTGTTTTCCCGCAACCGCAGACCCTGCTAGAGTTGCAA
CAGCG
CGAGCAGGCAGAGGCGGCGCTGCGAAAGGAAAGCTTCCGCAGGCCAAGCAGCTT
GTCCGA
TCTAGGCGCTGCGGCCCGCGGTCAGATGCTAGTAGCCATTTCCAAGCTTGATA
GGGTC
TCTTACCAGCACTCGCACCAACCGCCCGCGCCTGCTGGGCGAGGAGGAGTACCTA
AACAA
CTCGCTGCTGCAGCCGCAGCGCGAAAAAACCTGCCTCCGGCATTTCCTCAACAAC
GGGAT
AGAGAGCCTAGTGGACAAGATGAGTAGATGGAAGACGTACGCGCAGGAGCACAG
GGACGT
GCCAGGCCCGCGCCCGCCACCCGTCGTCAAAGGCACGACCGTCAGCGGGGTCTG
GTGTG
GGAGGACGATGACTCGGCAGACGACAGCAGCGTCCTGGATTTGGGAGGGAGTGG
CAACCC
GTTTGCACCTTCGCCCCAGGCTGGGGAGAATGTTTTAAAAAAAAAAAAAGCATG
ATGCA

AAATAAAAACTCACCAAGGCCATGGCACCGAGCGTTGGTTTTCTTGTATTCCCCT
TAGT
ATGCGGCGCGCGGCGATGTATGAGGAAGGTCCTCCTCCCTCCTACGAGAGTGTGG
TGAGC
GCGGCGCCAGTGGCGGCGGCGCTGGGTTCTCCCTTCGATGCTCCCCTGGACCCGC
CGTTT
GTGCCTCCGCGGTACCTGCGGCCTACCGGGGGGAGAAACAGCATCCGTTACTCTG
AGTTG
GCACCCCTATTTCGACACCACCCGTGTGTACCTGGTGGACAACAAGTCAACGGATG
TGGCA
TCCCTGAACTACCAGAACGACCACAGCAACTTTCTGACCACGGTCATTCAAAACA
ATGAC
TACAGCCCGGGGGAGGCAAGCACACAGACCATCAATCTTGACGACCGGTTCGCACT
GGGGC
GGCGACCTGAAAACCATCCTGCATACCAACATGCCAAATGTGAACGAGTTCATGT
TTACC
AATAAGTTTAAGGCGCGGGTGTGGTGTGCGCTTGCCTACTAAGGACAATCAGG
TGGAG
CTGAAATACGAGTGGGTGGAGTTCACGCTGCCCGAGGGCAACTACTCCGAGACCA
TGACC
ATAGACCTTATGAACAACGCGATCGTGGAGCACTACTTGAAAGTGGGCAGACAG
AACGGG
GTTCTGGAAAGCGACATCGGGGTAAAGTTTGACACCCGCAACTTCAGACTGGGGT
TTGAC
CCCGTCACTGGTCTTGTCATGCCTGGGGTATATACAAACGAAGCCTTCCATCCAGA
CATC
ATTTTGCTGCCAGGATGCGGGGTGGACTTCACCCACAGCCGCCTGAGCAACTTGT
TGGGC
ATCCGCAAGCGGCAACCCTTCCAGGAGGGCTTTAGGATCACCTACGATGATCTGG
AGGGT

GGTAACATTCCCGCACTGTTGGATGTGGACGCCTACCAGGCGAGCTTGAAAGATG
ACACC
GAACAGGGCGGGGGTGGCGCAGGCGGCAGCAACAGCAGTGGCAGCGGCGCGGA
AGAGAAC
TCCAACGCGGCAGCCGCGGCAATGCAGCCGGTGGAGGACATGAACGATCATGCC
ATTTCG
GGCGACACCTTTGCCACACGGGCTGAGGAGAAGCGCGCTGAGGCCGAAGCAGCG
GCCGAA
GCTGCCGCCCCGCTGCGCAACCCGAGGTGAGAAAGCCTCAGAAGAAACCGGTG
ATCAAA
CCCCTGACAGAGGACAGCAAGAAACGCAGTTACAACCTAATAAGCAATGACAGC
ACCTTC
ACCCAGTACCGCAGCTGGTACCTTGCATACTACGGCGACCCTCAGACCGGAA
TCCGC
TCATGGACCCTGCTTTGCACTCCTGACGTAACCTGCGGCTCGGAGCAGGTCTACTG
GTCG
TTGCCAGACATGATGCAAGACCCCGTGACCTTCCGCTCCACGCGCCAGATCAGCA
ACTTT
CCGGTGGTGGGCGCCGAGCTGTTGCCCGTGCACTCCAAGAGCTTCTACAACGACC
AGGCC
GTCTACTCCAACTCATCCGCCAGTTTACCTCTCTGACCCACGTGTTCAATCGCTTT
CCC
GAGAACCAGATTTTGGCGCGCCCGCCAGCCCCACCATCACCACCGTCAGTGAAA
ACGTT
CCTGCTCTCACAGATCACGGGACGCTACCGCTGCGCAACAGCATCGGAGGAGTCC
AGCGA
GTGACCATTACTGACGCCAGACGCCGCACCTGCCCTACGTTTACAAGGCCCTGG
GCATA
GTCTCGCCGCGGTCCTATCGAGCCGCACTTTTTGAGCAAGCATGTCCATCCTTAT
ATCG

CCCAGCAATAACACAGGCTGGGGCCTGCGCTTCCCAAGCAAGATGTTTGGCGGGG
CCAAG
AAGCGCTCCGACCAACACCCAGTGCGCGTGCGCGGGCACTACCGCGCGCCCTGGG
GCGCG
CACAAACGCGGCCGCACTGGGCGCACCACCGTCGATGACGCCATCGACGCGGTG
GTGGAG
GAGGCGGCAACTACACGCCACGCCGCCACCAGTGTCACAGTGGACGCGGCC
ATTCAG
ACCGTGGTGCGCGGAGCCCGGCGCTATGCTAAAATGAAGAGACGGCGGAGGCGC
GTAGCA
CGTCGCCACCGCCGCCGACCCGGCACTGCCGCCAACGCGCGGGCGGGCCCTGC
TTAAC
CGCGCACGTGCGACCCGGCCGACGGGCGGCCATGCGGGCCGCTCGAAGGCTGGCC
GCGGGT
ATTGTCACTGTGCCCCCAGGTCCAGGCGACGAGCGGCCGCCGACGAGCCGCGG
CCATT
AGTGCTATGACTCAGGGTCGCAGGGGCAACGTGTATTGGGTGCGCGACTCGGTTA
GCGGC
CTGCGCGTGCCCGTGCGCACCCGCCCCCGCGCAACTAGATTGCAAGAAAAAACT
ACTTA
GACTCGTACTGTTGTATGTATCCAGCGGCGGGCGCGCAACGAAGCTATGTCCA
AGCGC
AAAATCAAAGAAGAGATGCTCCAGGTCATCGCGCCGGAGATCTATGGCCCCCGA
AGAAG
GAAGAGCAGGATTACAAGCCCCGAAAGCTAAAGCGGGTCAAAAAGAAAAAGAA
AGATGAT
GATGATGAACTTGACGACGAGGTGGAAGTCTGACGCTACCGCGCCCAGGCGAC
GGGTA
CAGTGAAAGGTGCGACGCGTAAAACGTGTTTTGCGACCCGGCACCACCGTAGTCT
TTACG

CCCGGTGAGCGCTCCACCCGCACCTACAAGCGCGTGTATGATGAGGTGTACGGCG
ACGAG
GACCTGCTTGAGCAGGCCAACGAGCGCCTCGGGGAGTTTGCCTACGGAAAGCGGC
ATAAG
GACATGCTGGCGTTGCCGCTGGACGAGGGCAACCCAACACCTAGCCTAAAGCCCC
TAACA
CTGCAGCAGGTGCTGCCCCGCGCTTGCACCGTCCGAAGAAAAGCGCGGCCTAAAGC
GCGAG
TCTGGTGACTTGGCACCCACCGTGCAGCTGATGGTACCCAAGCGCCAGCGACTGG
AAGAT
GTCTTGAAAAAATGACCGTGGAACCTGGGCTGGAGCCCCGAGGTCCGCGTGCGGC
CAATC
AAGCAGGTGGCGCCGGGACTGGGCGTGCAGACCGTGGACGTTTCAGATACCCACT
ACCAGT
AGCACCAGTATTGCCACCGCCACAGAGGGCATGGAGACACAAACGTCCCCGGTTG
CCTCA
GCGGTGGCGGATGCCGCGGTGCAGGCGGTGCGTGCGGCCGCGTCCAAGACCTCTA
CGGAG
GTGCAAACGGACCCGTGGATGTTTCGCGTTTCAGCCCCCGGCGCCCGCGCGGTT
CGAGG
AAGTACGGCGCCGCCAGCGCGCTACTGCCGAATATGCCCTACATCCTTCCATTG
CGCCT
ACCCCCGGCTATCGTGGCTACACCTACCGCCCCAGAAGACGAGCAACTACCCGAC
GCCGA
ACCACCACTGGAACCCGCCGCGCCGTCGCCGTCGCCAGCCCGTGTGGCCCCGA
TTTCC
GTGCGCAGGGTGGCTCGCGAAGGAGGCAGGACCCTGGTGCTGCCAACAGCGCGC
TACCAC
CCCAGCATCGTTTAAAAGCCGGTCTTTGTGGTTCTTGCAGATATGGCCCTCACCTG
CCGC

CTCCGTTTCCCGGTGCCGGGATTCCGAGGAAGAATGCACCGTAGGAGGGGCATGG
CCGGC
CACGGCCTGACGGGCGGCATGCGTCGTGCGCACCAACCGGCGGCGGCGCGTCGC
ACCGT
CGCATGCGCGGCGGTATCCTGCCCTCCTTATTCCAATGATCGCCGCGGCGATTGG
CGCC
GTGCCCCGAATTGCATCCGTGGCCTTGCAGGCGCAGAGACACTGATTA AAAACAA
GTTGC
ATGTGGAAAAATCAAATAAAAAGTCTGGACTCTCACGCTCGCTTGGTCCTGTAA
CTATT
TTGTAGAATGGAAGACATCAACTTTGCGTCTCTGGCCCCGCGACACGGCTCGCGC
CCGTT
CATGGGAAACTGGCAAGATATCGGCACCAGCAATATGAGCGGTGGCGCCTTCAGC
TGGGG
CTCGCTGTGGAGCGGCATTA AAAATTTCCGGTCCACCGTTAAGAACTATGGCAGC
AAGGC
CTGGAACAGCAGCACAGGCCAGATGCTGAGGGATAAGTTGAAAGAGCAAAAATTT
CCAACA
AAAGGTGGTAGATGGCCTGGCCTCTGGCATTAGCGGGGTGGTGGACCTGGCCAAC
CAGGC
AGTGCAAAATAAGATTAACAGTAAGCTTGATCCCCGCCCTCCCGTAGAGGAGCCT
CCACC
GGCCGTGGAGACAGTGTCTCCAGAGGGGCGTGGCGAAAAGCGTCCGCGCCCCGA
CAGGGA
AGAAACTCTGGTGACGCAAATAGACGAGCCTCCCTCGTACGAGGAGGCACTAAA
GCAAGG
CCTGCCACCACCCGTCCCATCGCGCCCATGGCTACCGGAGTGCTGGGCCAGCAC
ACACC
CGTAACGCTGGACCTGCCTCCCCCGCCGACACCCAGCAGAAACCTGTGCTGCCA
GGCCC

GACCGCCGTTGTTGTAACCCGTCCTAGCCGCGCGTCCCTGCGCCGCGCCGCCAGC
GGTCC
GCGATCGTTGCGGCCCGTAGCCAGTGGCAACTGGCAAAGCACACTGAACAGCATC
GTGGG
TCTGGGGGTGCAATCCCTGAAGCGCCGACGATGCTTCTGAATAGCTAACGTGTGC
TATGT
GTGTCATGTATGCGTCCATGTCGCCGCCAGAGGAGCTGCTGAGCCGCCGCGCGCC
CGCTT
TCCAAGATGGCTACCCCTTCGATGATGCCGCAGTGGTCTTACATGCACATCTCGGG
CCAG
GACGCCTCGGAGTACCTGAGCCCCGGGCTGGTGCAGTTTGCCCGCGCCACCGAGA
CGTAC
TTCAGCCTGAATAACAAGTTTAGAAACCCACGGTGGCGCCTACGCACGACGTGA
CCACA
GACCGGTCCCAGCGTTTGACGCTGCGGTTTCATCCCTGTGGACCGTGAGGATACTG
CGTAC
TCGTACAAGGCGCGGTTACCCTAGCTGTGGGTGATAACCGTGTGCTGGACATGG
CTTCC
ACGTACTIONTTGACATCCGCGGCGTGCTGGACAGGGGCCCTACTTTTAAGCCCTACTC
TGGC
ACTGCCTACAACGCCCTGGCTCCCAAGGGTGCCCCAAATCCTTGCGAATGGGATG
AAGCT
GCTACTGCTCTTGAAATAAACCTAGAGAAGAGGACGATGACAACGAAGACGAA
GTAGAC
GAGCAAGCTGAGCAGCAAAAACTCACGTATTTGGGCAGGCGCCTTATTCTGGTA
TAAAT
ATTACAAAGGAGGGTATTCAAATAGGTGTCGAAGGTCAAACACCTAAATATGCCG
ATAAA
ACATTTCAACCTGAACCTCAAATAGGAGAATCTCAGTGGTACGAAACTGAAATTA
ATCAT

GCAGCTGGGAGAGTCCTTAAAAAGACTACCCCAATGAAACCATGTTACGGTTCAT
ATGCA
AAACCCACAAATGAAAATGGAGGGCAAGGCATTCTTGTAAGCAACAAAATGGA
AAGCTA
GAAAGTCAAGTGGAAATGCAATTTTTCTCAACTACTGAGGCGACCGCAGGCAATG
GTGAT
AACTTGACTCCTAAAGTGGTATTGTACAGTGAAGATGTAGATATAGAAACCCAG
ACACT
CATATTTCTTACATGCCCACTATTAAGGAAGGTAACCTCACGAGAACTAATGGGCC
AACAA
TCTATGCCCAACAGGCCTAATTACATTGCTTTTAGGGACAATTTTATTGGTCTAAT
GTAT
TACAACAGCACGGGTAATATGGGTGTTCTGGCGGGCCAAGCATCGCAGTTGAATG
CTGTT
GTAGATTTGCAAGACAGAAACACAGAGCTTTCATACCAGCTTTTGCTTGATTCCAT
TGGT
GATAGAACCAGGTACTTTTCTATGTGGAATCAGGCTGTTGACAGCTATGATCCAG
ATGTT
AGAATTATTGAAAATCATGGAAGTGAAGTGAAGTTCCAAATTACTGCTTTCCAC
TGGGA
GGTGTGATTAATACAGAGACTCTTACCAAGGTAAAACCTAAAACAGGTCAGGAA
AATGGA
TGGGAAAAAGATGCTACAGAATTTTCAGATAAAAATGAAATAAGAGTTGGAAAT
AATTTT
GCCATGGAAATCAATCTAAATGCCAACCTGTGGAGAAATTCCTGTACTCCAACA
TAGCG
CTGTATTTGCCCGACAAGCTAAAGTACAGTCCTTCCAACGTAAAAATTTCTGATAA
CCCA
AACACCTACGACTACATGAACAAGCGAGTGGTGGCTCCCGGGTTAGTGGACTGCT
ACATT

AACCTTGGAGCACGCTGGTCCCTTGACTATATGGACAACGTCAACCCATTTAACC
ACCAC
CGCAATGCTGGCCTGCGCTACCGCTCAATGTTGCTGGGCAATGGTCGCTATGTGCC
CTTC
CACATCCAGGTGCCTCAGAAGTTCTTTGCCATTA AAAACCTCCTTCTCCTGCCGGG
CTCA
TACACCTACGAGTGGA ACTTCAGGAAGGATGTTAACATGGTTCTGCAGAGCTCCC
TAGGA
AATGACCTAAGGGTTGACGGAGCCAGCATTAAAGTTTGATAGCATTTGCCTTTACG
CCACC
TTCTTCCCATGGCCCACAACACCGCCTCCACGCTTGAGGCCATGCTTAGAAAACG
ACACC
AACGACCAGTCCTTTAACGACTATCTCTCCGCCGCCAACATGCTCTACCCTATACC
CGCC
AACGCTACCAACGTGCCCATATCCATCCCCTCCCGCAACTGGGCGGCTTTCCGCG
GCTGG
GCCTTCACGCGCCTTAAGACTAAGGAAACCCCATCACTGGGCTCGGGCTACGACC
CTTAT
TACACCTACTCTGGCTCTATACCCTACCTAGATGGAACCTTTTACCTCAACCACAC
CTTT
AAGAAGGTGGCCATTACCTTTGACTCTTCTGTCAGCTGGCCTGGCAATGACCGCCT
GCTT
ACCCCAACGAGTTTGAAATTAAGCGCTCAGTTGACGGGGAGGGTTACAACGTTG
CCCAG
TGTAACATGACCAAAGACTGGTTCCTGGTACAAATGCTAGCTAACTACAACATTG
GCTAC
CAGGGCTTCTATATCCCAGAGAGCTACAAGGACCGCATGTACTCCTTCTTTAGAA
ACTTC
CAGCCCATGAGCCGTCAGGTGGTGGATGATACTAAATACAAGGACTACCAACAG
GTGGGC

ATCCTACACCAACACAACAACCTCTGGATTTGTTGGCTACCTTGCCCCACCATGCG
CGAA
GGACAGGCCTACCCTGCTAACTTCCCCTATCCGCTTATAGGCAAGACCGCAGTTG
ACAGC
ATTACCCAGAAAAAGTTTCTTTGCGATCGCACCTTTGGCGCATCCCATTCTCCAG
TAAC
TTTATGTCCATGGGCGCACTCACAGACCTGGGCCAAAACCTTCTCTACGCCAACTC
CGCC
CACGCGCTAGACATGACTTTTGAGGTGGATCCCATGGACGAGCCCACCCTTCTTTA
TGTT
TTGTTTGAAGTCTTTGACGTGGTCCGTGTGCACCGGCCGCACCGCGGCGTCATCGA
AACC
GTGTACCTGCGCACGCCCTTCTCGGCCGGAACGCCACAACATAAAGAAGCAAGC
AACAT
CAACAACAGCTGCCGCCATGGGCTCCAGTGAGCAGGAACTGAAAGCCATTGTCAA
AGATC
TTGGTTGTGGGCCATATTTTTTGGGCACCTATGACAAGCGCTTCCAGGCTTTGTT
TCTC
CACACAAGCTCGCCTGCGCCATAGTCAATACGGCCGGTCGCGAGACTGGGGGCGT
ACACT
GGATGGCCTTTGCCTGGAACCCGCACTCAAAAACATGCTACCTCTTTGAGCCCTTT
GGCT
TTTCTGACCAGCGACTCAAGCAGGTTTACCAGTTTGGAGTACGAGTCACTCCTGCGC
CGTA
GCGCCATTGCTTCTTCCCCGACCGCTGTATAACGCTGGAAAAGTCCACCCAAAG
CGTAC
AGGGGCCCAAACCTCGGCCGCTGTGGACTATTCTGCTGCATGTTTCTCCACGCCTTT
GCCA
ACTGGCCCCAAACCTCCCATGGATCACAACCCCAACCATGAACCTTATTACCGGGGT
ACCCA

ACTCCATGCTCAACAGTCCCCAGGTACAGCCCACCCTGCGTCGCAACCAGGAACA
GCTCT
ACAGCTTCCTGGAGCGCCACTCGCCCTACTTCCGCAGCCACAGTGCGCAGATTAG
GAGCG
CCACTTCTTTTTGTCACTTGAAAAACATGTAAAAATAATTACTTATGACTCGTACT
ATTG
TTATTCATCCAGGCGGTAGGAGGGCCATCATGGCTATGATGGAGGTCCAGGGGGG
ACCCA
GCCTGGGACAGACCTGCGTGCTGATCGTGATCTTTACAGTGCTCCTGCAGTCTCTC
TGTG
TGGCTGTAACCTTACGTGTACTTTACCAACGAGCTGAAGCAGATGCAGGACAAGTA
CTCCA
AAAGTGGCATTGCTTGTTTCTTAAAAGAAGATGACAGTTATTGGGACCCCAATGA
CGAAG
AGAGTATGAACAGCCCCTGCTGGCAAGTCAAGTGGCAACTCCGTCAGCTCGTTAG
AAAGA
TGATTTTGAGAACCTCTGAGGAAACCATTTCTACAGTTCAAGAAAAGCAACAAAA
TATTT
CTCCCCTAGTGAGAGAAAGAGGTCCTCAGAGAGTAGCAGCTCACATAACTGGGAC
CAGAG
GAAGAAGCAACACATTGTCTTCTCCAACTCCAAGAATGAAAAGGCTCTGGGCCG
CAAAA
TAAACTCCTGGGAATCATCAAGGAGTGGGCATTCATTCCTGAGCAACTTGCACTT
GAGGA
ATGGTGAACCTGGTCATCCATGAAAAAGGGTTTTACTACATCTATTCCCAAACATA
CTTTC
GATTCAGGAGGAAATAAAAAGAAAACACAAAGAACGACAAACAAATGGTCCAAT
ATATTT
ACAAATACACAAGTTATCCTGACCCTATATTGTTGATGAAAAGTGCTAGAAATAG
TTGTT

GGTCTAAAGATGCAGAATATGGACTCTATTCCATCTATCAAGGGGGAATATTTGA
GCTTA
AGGAAAATGACAGAATTTTTGTTTCTGTAACAAATGAGCACTTAATAGACATGGA
CCATG
AAGCCAGTTTTTTTCGGGGCCTTTTTAGTTGGCTAAGCTAGCTACTAGAGACACTTT
CAAT
AAAGGCAAATGCTTTTATTTGTACACTCTCGGGTGATTATTTACCCCCACCCTTGC
CGTC
TGCGCCGTTTAAAAATCAAAGGGGTCTGCCGCGCATCGCTATGCGCCACTGGCA
GGGAC
ACGTTGCGATACTGGTGTTTAGTGCTCCACTTAAACTCAGGCACAACCATCCGCG
GCAGC
TCGGTGAAGTTTTCACTCCACAGGCTGCGCACCATCACCAACGCGTTTAGCAGGT
CGGGC
GCCGATATCTTGAAGTCGCAGTTGGGGCCTCCGCCCTGCGCGCGAGTTGCGAT
ACACA
GGGTTGCAGCACTGGAACACTATCAGCGCCGGGTGGTGCACGCTGGCCAGCACGC
TCTTG
TCGGAGATCAGATCCGCGTCCAGGTCCTCCGCGTTGCTCAGGGCGAACGGAGTCA
ACTTT
GGTAGCTGCCTTCCCAAAAAGGGCGCGTGCCAGGCTTTGAGTTGCACTCGCACC
GTAGT
GGCATCAAAGGTGACCGTGCCCGGTCTGGGCGTTAGGATACAGCGCCTGCATAA
AAGCC
TTGATCTGCTTAAAAGCCACCTGAGCCTTTGCGCCTTCAGAGAAGAACATGCCGC
AAGAC
TTGCCGAAAACCTGATTGGCCGGACAGGCCGCGTCGTGCACGCAGCACCTTGCCT
CGGTG
TTGGAGATCTGCACCACATTTTCGGCCCCACCGGTTCTTCACGATCTTGGCCTTGCT
AGAC

TGCTCCTTCAGCGCGCGCTGCCCGTTTTTCGCTCGTCACATCCATTTCAATCACGTG
CTCC
TTATTTATCATAATGCTTCCGTGTAGACACTTAAGCTCGCCTTCGATCTCAGCGCA
GCGG
TGCAGCCACAACGCGCAGCCCGTGGGCTCGTGATGCTTGTAGGTCACCTCTGCAA
ACGAC
TGCAGGTACGCCTGCAGGAATCGCCCCATCATCGTCACAAAGGTCTTGTTGCTGG
TGAAG
GTCAGCTGCAACCCGCGGTGCTCCTCGTTCAGCCAGGTCTTGCATACGGCCGCCA
GAGCT
TCCACTTGGTCAGGCAGTAGTTTGAAGTTCGCCTTTAGATCGTTATCCACGTGGTA
CTTG
TCCATCAGCGCGCGCGCAGCCTCCATGCCCTTCTCCCACGCAGACACGATCGGCA
CACTC
AGCGGGTTCATCACCGTAATTTCACTTTCCGCTTCGCTGGGCTCTTCCTCTTCCTCT
TGC
GTCCGCATACCACGCGCCACTGGGTGCTCTTCATTCAGCCGCCGCACTGTGCGCTT
ACCT
CCTTTGCCATGCTTGATTAGCACCGGTGGGTTGCTGAAACCCACCATTTGTAGCGC
CACA
TCTTCTCTTTCTTCCTCGCTGTCCACGATTACCTCTGGTGATGGCGGGCGCTCGGG
CTTG
GGAGAAGGGCGCTTCTTTTTCTTCTTGGGCGCAATGGCCAAATCCGCCGCCGAGG
TCGAT
GGCCGCGGGCTGGGTGTGCGCGGCACCAGCGCGTCTTGTGATGAGTCTTCCTCGT
CCTCG
GACTCGATACCCGCCTCATCCGCTTTTTTGGGGGCGCCCGGGGAGGCGGGCGGG
ACGGG
GACGGGGACGACACGTCCTCCATGGTTGGGGGACGTGCGCGCCGACCCGCGTCCGC
GCTCG

GGGGTGGTTTCGCGCTGCTCCTCTTCCCGACTGGCCATTTCTTCTCCTATAGGCA
GAAA
AAGATCATGGAGTCAGTCGAGAAGAAGGACAGCCTAACCGCCCCCTCTGAGTTCG
CCACC
ACCGCCTCCACCGATGCCGCCAACGCGCCTACCACCTTCCCCGTCGAGGCACCCC
CGCTT
GAGGAGGAGGAAGTGATTATCGAGCAGGACCCAGGTTTTGTAAGCGAAGACGAC
GAGGAC
CGCTCAGTACCAACAGAGGATAAAAAGCAAGACCAGGACAACGCAGAGGCAAAC
GAGGAA
CAAGTCGGGCGGGGGGACGAAAGGCATGGCGACTACCTAGATGTGGGAGACGAC
GTGCTG
TTGAAGCATCTGCAGCGCCAGTGCGCCATTATCTGCGACGCGTTGCAAGAGCGCA
GCGAT
GTGCCCTCGCCATAGCGGATGTCAGCCTTGCCCTACGAACGCCACCTATTCTCACC
GCGC
GTACCCCCAAACGCCAAGAAAACGGCACATGCGAGCCCAACCCGCGCCTCAACT
TCTAC
CCCGTATTTGCCGTGCCAGAGGTGCTTGCCACCTATCACATCTTTTTCCAAAAGT
CAAG
ATACCCCTATCCTGCCGTGCCAACCGCAGCCGAGCGGACAAGCAGCTGGCCTTGC
GGCAG
GGCGCTGTCATACCTGATATCGCCTCGCTCAACGAAGTGCCAAAAATCTTTGAGG
GTCTT
GGACGCGACGAGAAGCGCGCGGCAAACGCTCTGCAACAGGAAAACAGCGAAAAT
GAAAGT
CACTCTGGAGTGTTGGTGGAACTCGAGGGTGACAACGCGCGCCTAGCCGTACTAA
AACGC
AGCATCGAGGTCACCCACTTTGCCTACCCGGCACTTAACCTACCCCCAAGGTCAT
GAGC

ACAGTCATGAGTGAGCTGATCGTGCGCCGTGCGCAGCCCCTGGAGAGGGATGCAA
ATTTG
CAAGAACAAACAGAGGAGGGCCTACCCGCAGTTGGCGACGAGCAGCTAGCGCGC
TGGCTT
CAAACGCGCGAGCCTGCCGACTTGGAGGAGCGACGCAAATAATGATGGCCGCA
GTGCTC
GTTACCGTGGAGCTTGAGTGCATGCAGCGGTTCTTTGCTGACCCGGAGATGCAGC
GCAAG
CTAGAGGAAACATTGCACTACACCTTTCGACAGGGCTACGTACGCCAGGCCTGCA
AGATC
TCCAACGTGGAGCTCTGCAACCTGGTCTCCTACCTTGAATTTGCACGAAAACCG
CCTT
GGGCAAAACGTGCTTCATTCCACGCTCAAGGGCGAGGCGCGCCGCGACTACGTCC
GCGAC
TGCGTTTACTTATTTCTATGCTACACCTGGCAGACGGCCATGGGCGTTTGGCAGCA
GTGC
TTGGAGGAGTGCAACCTCAAGGAGCTGCAGAACTGCTAAAGCAAACTGAAG
GACCTA
TGGACGGCCTTCAACGAGCGCTCCGTGGCCGCGCACCTGGCGGACATCATTTTCC
CCGAA
CGCCTGCTTAAAACCCTGCAACAGGGTCTGCCAGACTTCACCAGTCAAAGCATGT
TGCAG
AACTTTAGGAACTTTATCCTAGAGCGCTCAGGAATCTTGCCCGCCACCTGCTGTGC
ACTT
CCTAGCGACTTTGTGCCATTAAGTACCGCGAATGCCCTCCGCCGCTTTGGGGCCA
CTGC
TACCTTCTGCAGCTAGCCAACTACCTTGCCCTACCACTCTGACATAATGGAAGACGT
GAGC
GGTGACGGTCTACTGGAGTGTCAGTGTGCGCTGCAACCTATGCACCCCGCACCGCT
CCCTG

GTTTGCAATTCGCAGCTGCTTAACGAAAGTCAAATTATCGGTACCTTTGAGCTGCA
GGGT
CCCTCGCCTGACGAAAAGTCCGCGGCTCCGGGGTTGAAACTCACTCCGGGGCTGT
GGACG
TCGGCTTACCTTCGCAAATTTGTACCTGAGGACTACCACGCCACGAGATTAGGTT
CTAC
GAAGACCAATCCCGCCCGCAAATGCGGAGCTTACCGCCTGCGTCATTACCCAGG
GCCAC
ATTCTTGGCCAATTGCAAGCCATCAACAAAGCCCGCCAAGAGTTTCTGCTACGAA
AGGGA
CGGGGGGTTTACTTGGACCCCCAGTCCGGCGAGGAGCTCAACCCAATCCCCCGC
CGCCG
CAGCCCTATCAGCAGCAGCCGCGGGCCCTTGCTTCCCAGGATGGCACCCAAAAAG
AAGCT
GCAGCTGCCGCCGCCACCCACGGACGAGGAGGAATACTGGGACAGTCAGGCAGA
GGAGGT
TTTGGACGAGGAGGAGGAGGACATGATGGAAGACTGGGAGAGCCTAGACGAGGA
AGCTTC
CGAGGTGGAAGAGGTGTCAGACGAAACACCGTCAACCTCGGTTCGCATTCCCCTCG
CCGGC
GCCCCAGAAATCGGCAACCGGTTCCAGCATGGCTACAACCTCCGCTCCTCAGGCG
CCGCC
GGCACTGCCCGTTCGCCGACCCAACCGTAGATGGGACACCACTGGAACCAGGGCC
GGTAA
GTCCAAGCAGCCGCCGCGTTAGCCCAAGAGCAACAACAGCGCCAAGGCTACCG
CTCATG
GCGCGGGCACAAGAACGCCATAGTTGCTTGCTTGCAAGACTGTGGGGGCAACATC
TCCTT
CGCCCCGCGCTTTCTTCTCTACCATCACGGCGTGGCCTTCCCCGTAACATCCTGC
ATTA

CTACCGTCATCTCTACAGCCCATACTGCACCGGCGGCAGCGGCAGCGGCAGCAAC
AGCAG
CGGCCACACAGAAGCAAAGGCGACCGGATAGCAAGACTCTGACAAAGCCCAAGA
AATCCA
CAGCGGCGGCAGCAGCAGGAGGAGGAGCGCTGCGTCTGGCGCCCAACGAACCCG
TATCGA
CCCGCGAGCTTAGAAACAGGATTTTTCCCACTCTGTATGCTATATTTCAACAGAGC
AGGG
GCCAAGAACAAGAGCTGAAAATAAAAAACAGGTCTCTGCGATCCCTCACCCGCA
GCTGCC
TGTATCACAAAAGCGAAGATCAGCTTCGGCGCACGCTGGAAGACGCGGAGGCTCT
CTTCA
GTAAATACTGCGCGCTGACTCTTAAGGACTAGTTTCGCGCCCTTTCTCAAATTTAA
GCGC
GAAAACACTACGTCATCTCCAGCGGCCACACCCGGCGCCAGCACCTGTCGTCAGCGC
CATT
TGAGCAAGGAAATTCCCACGCCCTACATGTGGAGTTACCAGCCACAAATGGGACT
TGCGG
CTGGAGCTGCCCAAGACTACTCAACCCGAATAAACTACATGAGCGCGGGACCCCA
CATGA
TATCCCGGGTCAACGGAATCCGCGCCCACCGAAACCGAATTCTCTTGGAACAGGC
GGCTA
TTACCACCACCTCGTAATAACCTTAATCCCGTAGTTGGCCCGCTGCCCTGGTG
TACC
AGGAAAGTCCCGCTCCCACCACTGTGGTACTTCCCAGAGACGCCAGGCCGAAGT
TCAGA
TGAATAACTCAGGGGCGCAGCTTGCGGGCGGCTTTCGTACAGGGTGCGGTGCC
CGGGC
AGGGTATAACTCACCTGACAATCAGAGGGCGAGGTATTCAGCTCAACGACGAGTC
GGTGA

GCTCCTCGCTTGGTCTCCGTCCGGACGGGACATTCAGATCGGGCGGCGCCGGCCG
TCCTT
CATTACGCCTCGTCAGGCAATCCTAACTCTGCAGACCTCGTCCTCTGAGCCGCGC
TCTG
GAGGCATTGGA ACTCTGCAATTTATTGAGGAGTTTGTGCCATCGGTCTACTTTAAC
CCCT
TCTCGGGACCTCCCGGCCACTATCCGGATCAATTTATTCCTAACTTTGACGCGGTA
AAGG
ACTCGGCGGACGGCTACGACTGAATGTTAAGTGGAGAGGCAGAGCAACTGCGCC
TGAAAC
ACCTGGTCCACTGTCGCCGCCACAAGTGCTTTGCCCGCGACTCCGGTGAGTTTTGC
TACT
TTGAATTGCCCGAGGATCATATCGAGGGCCCGGCGCACGGCGTCCGGCTTACCGC
CCAGG
GAGAGCTTGCCCGTAGCCTGATTCGGGAGTTTACCCAGCGCCCCCTGCTAGTTGA
GCGGG
ACAGGGGACCCTGTGTTCTCACTGTGATTTGCAACTGTCCTAACCTTGGATTACAT
CAAG
ATCTTTGTTGCCATCTCTGTGCTGAGTATAATAAATACAGAAATTTAAAATATACTG
GGGC
TCCTATCGCCATCCTGTAAACGCCACCGTCTTCACCCGCCCAAGCAAACCAAGGC
GAACC
TTACCTGGTACTTTTAACATCTCTCCCTCTGTGATTTACAACAGTTTCAACCCAGA
CGGA
GTGAGTCTACGAGAGAACCTCTCCGAGCTCAGCTACTCCATCAGAAAAACACCA
CCCTC
CTTACCTGCCGGGAACGTACGAGTGCGTCAACCGGCCGCTGCACCACACCTACCGC
CTGAC
CGTAAACCAGACTTTTTCCGGACAGACCTCAATAACTCTGTTTACCAGAACAGGA
GGTGA

GCTTAGAAAACCCCTAGGGTATTAGGCCAAAGGCGCAGCTACTGTGGGGTTTATG
AACAA
TTCAAGCAACTCTACGGGCTATTCTAATTCAGGTTTCTCTAGAATCGGGGTGGGG
TTAT
TCTCTGTCTTGATTCCTTTATTCTTATACTAACGCTTCTCTGCCTAAGGCTCGC
CGC
CTGCTGTGTGCACATTTGCATTTATTGTCAGCTTTTTAAACGCTGGGGTCGCCACC
CAAG
ATGATTAGGTACATAATCCTAGGTTTACTCACCCCTTGCCTCAGCCCACGGTACCAC
CCAA
AAGGTGGATTTTAAGGAGCCAGCCTGTAATGTTACATTCGCAGCTGAAGCTAATG
AGTGC
ACCACTTTATAAAATGCACCACAGAACATGAAAAGCTGCTTATTCGCCACAAAA
ACAAA
ATTGGCAAGTATGCTGTTTATGCTATTTGGCAGCCAGGTGACACTACAGAGTATA
ATGTT
ACAGTTTTCCAGGGTAAAAGTCATAAACTTTTATGTATACTTTTCCATTTTATGA
AATG
TGCGACATTACCATGTACATGAGCAAACAGTATAAGTTGTGGCCCCACAAAATT
GTGTG
GAAAACACTGGCACTTTCTGCTGCACTGCTATGCTAATTACAGTGCTCGCTTTGGT
CTGT
ACCCTACTCTATATTAATAACAAAAGCAGACGCAGCTTTATTGAGGAAAAGAAAA
TGCCT
TAATTTACTAAGTTACAAAGCTAATGTCACCACTAACTGCTTTACTCGCTGCTTGC
AAAA
CAAATTCAAAAGTTAGCATTATAATTAGAATAGGATTTAAACCCCCGGTCATT
TCCTG
CTCAATACCATTCCCCTGAACAATTGACTCTATGTGGGATATGCTCCAGCGCTACA
ACCT

TGAAGTCAGGCTTCCTGGATGTCAGCATCTGACTTTGGCCAGCACCTGTCCCGCGG
ATTT
GTTCCAGTCCAACACTACAGCGACCCACCCTAACAGAGATGACCAACACAACCAACG
CGGCC
GCCGCTACCGGACTTACATCTACCACAAATACACCCCAAGTTTCTGCCTTTGTCAA
TAAC
TGGGATAACTTGGGCATGTGGTGGTTCTCCATAGCGCTTATGTTTGTATGCCTTAT
TATT
ATGTGGCTCATCTGCTGCCTAAAGCGCAAACGCGCCCGACCACCCATCTATAGTC
CCATC
ATTGTGCTACACCCAAACAATGATGGAATCCATAGATTGGACGGACTGAAACACA
TGTTT
TTTTCTCTTACAGTATGATTAATGAGACATGATTCCTCGAGTTTTTATATTACTGA
CCC
TTGTTGCGCTTTTTTGTGCGTGCTCCACATTGGCTGCGGTTTCTCACATCGAAGTA
GACT
GCATTCCAGCCTTCACAGTCTATTTGCTTTACGGATTTGTCACCCTCACGCTCATCT
GCA
GCCTCATCACTGTGGTCATCGCCTTTATCCAGTGCATTGACTGGGTCTGTGTGCGC
TTTG
CATATCTCAGACACCATCCCCAGTACAGGGACAGGACTATAGCTGAGCTTCTTAG
AATTC
TTTAATTATGAAATTTACTGTGACTTTTCTGCTGATTATTTGCACCCTATCTGCGTT
TTG
TTCCCCGACCTCCAAGCCTCAAAGACATATATCATGCAGATTCACCTCGTATATGGA
ATAT
TCCAAGTTGCTACAATGAAAAAAGCGATCTTTCCGAAGCCTGGTTATATGCAATC
ATCTC
TGTTATGGTGTCTGCAGTACCATCTTAGCCCTAGCTATATATCCCTACCTTGACA
TTGG

CTGGAACGCAATAGATGCCATGAACCACCCAACCTTTCCCCGCGCCCGCTATGCTT
CCACT
GCAACAAGTTGTTGCCGGCGGCTTTGTCCCAGCCAATCAGCCTCGCCACCTTCTC
CCAC
CCCCACTGAAATCAGCTACTTTAATCTAACAGGAGGAGATGACTGACACCCTAGA
TCTAG
AAATGGACGGAATTATTACAGAGCAGCGCCTGCTAGAAAGACGCAGGGCAGCGG
CCGAGC
AACAGCGCATGAATCAAGAGCTCCAAGACATGGTTAACTTGCACCAGTGCAAAA
GGGGTA
TCTTTTGTCTGGTAAAGCAGGCCAAAGTCACCTACGACAGTAATACCACCGGACA
CCGCC
TTAGCTACAAGTTGCCAACCAAGCGTCAGAAATTGGTGGTCATGGTGGGAGAAAA
GCCCA
TTACCATAACTCAGCACTCGGTAGAAACCGAAGGCTGCATTCACTCACCTTGTC
AGGAC
CTGAGGATCTCTGCACCCTTATTAAGACCCTGTGCGGTCTCAAAGATCTTATTCCC
TTTA
ACTAATAAAAAAAAAATAATAAAGCATCACTTACTTAAAATCAGTTAGCAAATTC
TGTCC
AGTTTATTCAGCAGCACCTCCTTGCCCTCCTCCCAGCTCTGGTATTGCAGCTTCCTC
CTG
GCTGCAAACCTTTCTCCACAATCTAAATGGAATGTCAGTTTCCTCCTGTTCCCTGTCC
ATCC
GCACCCACTATCTTCATGTTGTTGCAGATGAAGCGCGCAAGACCGTCTGAAGATA
CCTTC
AACCCCGTGTATCCATATGACACGGAAACCGGTCTCCAACCTGTGCCTTTTCTTAC
TCCT
CCCTTTGTATCCCCAATGGGTTTCAAGAGAGTCCCCTGGGGTACTCTCTTTGCG
CCTA

TCCGAACCTCTAGTTACCTCCAATGGCATGCTTGCGCTCAAAATGGGCAACGGCC
TCTCT
CTGGACGAGGCCGGCAACCTTACCTCCCAAAATGTAACCACTGTGAGCCCACCTC
TCAA
AAAACCAAGTCAAACATAAACCTGGAAATATCTGCACCCCTCACAGTTACCTCAG
AAGCC
CTAACTGTGGCTGCCGCCGCACCTCTAATGGTCGCGGGCAACACACTCACCATGC
AATCA
CAGGCCCGCTAACCGTGCACGACTCCAACTTAGCATTGCCACCCAAGGACCCC
TCACA
GTGTCAGAAGGAAAGCTAGCCCTGCAAACATCAGGCCCCCTCACCACCACCGATA
GCAGT
ACCCTTACTATCACTGCCTCACCCCTCTAACTACTGCCACTGGTAGCTTGGGCAT
TGAC
TTGAAAGAGCCCATTTATACACAAAATGGAAAAGTAGGACTAAAGTACGGGGCTC
CTTTG
CATGTAACAGACGACCTAAACACTTTGACCGTAGCAACTGGTCCAGGTGTGACTA
TTAAT
AATACTTCCTTGCAAACCTAAAGTTACTGGAGCCTTGGGTTTTGATTCACAAGGCAA
TATG
CAACTTAATGTAGCAGGAGGACTAAGGATTGATTCTCAAAACAGACGCCTTATAC
TTGAT
GTTAGTTATCCGTTTGATGCTCAAAACCAACTAAATCTAAGACTAGGACAGGGCC
CTCTT
TTTATAAACTCAGCCCACAACCTGGATATTAAC TACAACAAAGGCCTTTACTTGTT
TACA
GCTTCAAACAATTCCAAAAAGCTTGAGGTAAACCTAAGCACTGCCAAGGGGTTGA
TGTTT
GACGCTACAGCCATAGCCATTAATGCAGGAGATGGGCTTGAATTTGGTTCACCTA
ATGCA

CCAAACACAAATCCCCTCAAACAAAAATTGGCCATGGCCTAGAATTTGATTCAA
ACAAG
GCTATGGTTCCTAAACTAGGAACTGGCCTTAGTTTTGACAGCACAGGTGCCATTAC
AGTA
GGAAACAAAAATAATGATAAGCTAACTTTGTGGACCACACCAGCTCCATCTCCTA
ACTGT
AGACTAAATGCAGAGAAAGATGCTAAACTCACTTTGGTCTTAACAAAATGTGGCA
GTCAA
ATACTTGCTACAGTTTCAGTTTTGGCTGTTAAAGGCAGTTTGGCTCCAATATCTGG
AACA
GTTCAAAGTGCTCATCTTATTATAAGATTTGACGAAAATGGAGTGCTACTAAACA
ATTCC
TTCCTGGACCCAGAATATTGGAACTTTAGAAATGGAGATCTTACTGAAGGCACAG
CCTAT
ACAAACGCTGTTGGATTTATGCCTAACCTATCAGCTTATCCAAAATCTCACGGTAA
AACT
GCCAAAAGTAACATTGTCAGTCAAGTTTACTTAAACGGAGACAAAACCTAAACCTG
TAACA
CTAACCATTACACTAAACGGTACACAGGAAACAGGAGACACAACCTCCAAGTGCA
TACTCT
ATGTCATTTTCATGGGACTGGTCTGGCCACAACCTACATTAATGAAATATTTGCCAC
ATCC
TCTTACACTTTTTCATACATTGCCCAAGAATAAAGAAATCGTTTGTGTTATGTTTCA
ACGT
GTTTATTTTTCAATTGCAGAAAATTTCAAGTCATTTTTCATTTCAGTAGTATAGCCCC
ACC
ACCACATAGCTTATACAGATCACCGTACCTTAATCAAACCTCACAGAACCCTAGTA
TTCAA
CCTGCCACCTCCCTCCCAACACACAGAGTACACAGTCCTTTCTCCCCGGCTGGCCT
TAAA

AAGCATCATATCATGGGTAACAGACATATTCTTAGGTGTTATATTCCACACGGTTT
CCTG
TCGAGCCAAACGCTCATCAGTGATATTAATAAACTCCCCGGGCAGCTCACTTAAG
TTCAT
GTCGCTGTCCAGCTGCTGAGCCACAGGCTGCTGTCCAACCTGCGGTTGCTTAACGG
GCGG
CGAAGGAGAAGTCCACGCCTACATGGGGGTAGAGTCATAATCGTGCATCAGGAT
AGGGCG
GTGGTGCTGCAGCAGCGCGGAATAAACTGCTGCCGCCGCGCTCCGTCCTGCAG
GAATA
CAACATGGCAGTGGTCTCCTCAGCGATGATTCGCACCGCCCGCAGCATAAGGCGC
CTTGT
CCTCCGGGCACAGCAGCGCACCTGATCTCACTTAAATCAGCACAGTAACTGCAG
CACAG
CACCACAATATTGTTCAAAATCCCACAGTGCAAGGCGCTGTATCCAAAGCTCATG
GCGGG
GACCACAGAACCCACGTGGCCATCATAACCACAAGCGCAGGTAGATTAAGTGGCG
ACCCCT
CATAAACACGCTGGACATAAACATTACCTCTTTTGGCATGTTGTAATTCACCACCT
CCCG
GTACCATATAAACCTCTGATTAACATGGCGCCATCCACCACCATCCTAAACCAG
CTGGC
CAAAACCTGCCCGCCGGCTATACACTGCAGGGAACCGGGACTGGAACAATGACA
GTGGAG
AGCCCAGGACTCGTAACCATGGATCATCATGCTCGTCATGATATCAATGTTGGCA
CAACA
CAGGCACACGTGCATACACTTCCTCAGGATTACAAGCTCCTCCCGGTTAGAACC
ATATC
CCAGGGAACAACCCATTCCTGAATCAGCGTAAATCCCACACTGCAGGGAAGACCT
CGCAC

GTA ACTCACGTTGTGCATTGTCAAAGTGTTACATTCGGGCAGCAGCGGATGATCC
TCCAG
TATGGTAGCGCGGGTTTCTGTCTCAAAGGAGGTAGACGATCCCTACTGTACGGA
GTGCG
CCGAGACAACCGAGATCGTGTGGTCGTAGTGTCATGCCAAATGGAACGCCGGAC
GTAGT
CATATTTCTGAAGCAAACCAGGTGCGGGCGTGACAAACAGATCTGCGTCTCCG
GTCTC
GCCGCTTAGATCGCTCTGTGTAGTAGTTGTAGTATATCCACTCTCTCAAAGCATCC
AGGC
GCCCCCTGGCTTCGGGTTCTATGTAAACTCCTTCATGCGCCGCTGCCCTGATAACA
TCCA
CCACCGCAGAATAAGCCACACCCAGCCAACCTACACATTCGTTCTGCGAGTCACA
CACGG
GAGGAGCGGGAAGAGCTGGAAGAACCATGTTTTTTTTTTATTCCAAAAGATTAT
CCAAA
ACCTCAAAATGAAGATCTATTAAGTGAACGCGCTCCCCTCCGGTGGCGTGGTCAA
ACTCT
ACAGCCAAAGAACAGATAATGGCATTGTGTAAGATGTTGCACAATGGCTTCCAAAA
GGCAA
ACGGCCCTCACGTCCAAGTGGACGTAAAGGCTAAACCCTTCAGGGTGAATCTCCT
CTATA
AACATTCCAGCACCTTCAACCATGCCCAAATAATTCTCATCTCGCCACCTTCTCAA
TATA
TCTCTAAGCAAATCCCGAATATTAAGTCCGGCCATTGTAAAAATCTGCTCCAGAG
CGCCC
TCCACCTTCAGCCTCAAGCAGCGAATCATGATTGCAAAAATTCAGGTTCTCACA
GACCT
GTATAAGATTCAAAAGCGGAACATTAACAAAAATACCGCGATCCCGTAGGTCCCT
TCGCA

GGGCCAGCTGAACATAATCGTGCAGGTCTGCACGGACCAGCGCGGCCACTTCCCC
GCCAG
GAACCTTGACAAAAGAACCCACACTGATTATGACACGCATACTCGGAGCTATGCT
AACCA
GCGTAGCCCCGATGTAAGCTTTGTTGCATGGGCGGCGATATAAAATGCAAGGTGC
TGCTC
AAAAAATCAGGCAAAGCCTCGCGCAAAAAAGAAAGCACATCGTAGTCATGCTCA
TGCAGA
TAAAGGCAGGTAAGCTCCGGAACCACCACAGAAAAAGACACCATTTTTCTCTCAA
ACATG
TCTGCGGGTTTCTGCATAAACACAAAATAAAATAACAAAAAACATTTAAACATT
AGAAG
CCTGTCTTACAACAGGAAAAACAACCCTTATAAGCATAAGACGGACTACGGCCAT
GCCGG
CGTGACCGTAAAAAACTGGTCACCGTGATTA AAAAGCACCACCGACAGCTCCTC
GGTCA
TGTCCGGAGTCATAATGTAAGACTCGGTAAACACATCAGGTTGATTCATCGGTCA
GTGCT
AAAAAGCGACCGAAATAGCCCGGGGAATACATACCCGCAGGCGTAGAGACAAC
ATTACA
GCCCCATAGGAGGTATAACAAAATTAATAGGAGAGAAAAACACATAAACACCT
GAAAAA
CCCTCCTGCCTAGGCAAAATAGCACCCCTCCCGCTCCAGAACAACATACAGCGCTT
CCACA
GCGGCAGCCATAACAGTCAGCCTTACCAGTAAAAAAGAAAACCTATTA AAAAAA
CACCAC
TCGACACGGCACCAGCTCAATCAGTCACAGTGTA AAAAAAGGGCCAAGTGCAGAG
CGAGTA
TATATAGGACTAAAAAATGACGTAACGGTTAAAGTCCACAAAAACACCCAGAA
AACCGC

ACGCGAACCTACGCCAGAAACGAAAGCCAAAAACCCACAACCTTCCTCAAATC
GTCACT

TCCGTTTTCCACGTTACGTCACCTCCCATTTTAATTAAGAAAACACTACAATTCCCA
ACAC

ATACAAGTTACTCCGCCCTAAACCTACGTCACCCGCCCGTTCCACGCCCGCG
CCAC

GTCACAAACTCCACCCCTCATTATCATATTGGCTTCAATCCAAAATAAGGTATAT
TATT

GATGATGATTACCCT

SEQ ID NO:52 是寡核苷酸引物 1618.95.1 的序列

5' TACCGGGGTACCCAACCTCCA

SEQ ID NO:53 是寡核苷酸引物 1618.95.6 的序列

5' GACGCGGCCTGTCCGGCC

SEQ ID NO:54 是寡核苷酸引物 1618.97.1 的序列

5' TACTTATGACTCGTACTATTGTTATTCATCCAGGCGGTAGGAGGGCCATC
ATGAA

SEQ ID NO: 55 是寡核苷酸引物 1618.97.2 的序列

5' CCTTTATTGAAAGTGTCTCTAGTAGCTAGCGGGAGGGAGGTCC

SEQ ID NO: 56 是寡核苷酸引物 1618.95.5 的序列

5' TACTAGAGACACTTTCAATAAAGG

SEQ ID NO: 57 是寡核苷酸引物 1706.83.1 的序列

5' GTT AAC ATG GTT CTG CAG AGC

SEQ ID NO: 58 是寡核苷酸引物 1706.83.2 的序列

5' GGC TCG TCC ATG GGA TCC ACC TCA AAA GTC

SEQ ID NO: 59 是寡核苷酸引物 1706.95.1 的序列

5' GGA TCC CAT GGA CGA GCC CA

SEQ ID NO: 60 是寡核苷酸引物 1618.116.3 的序列

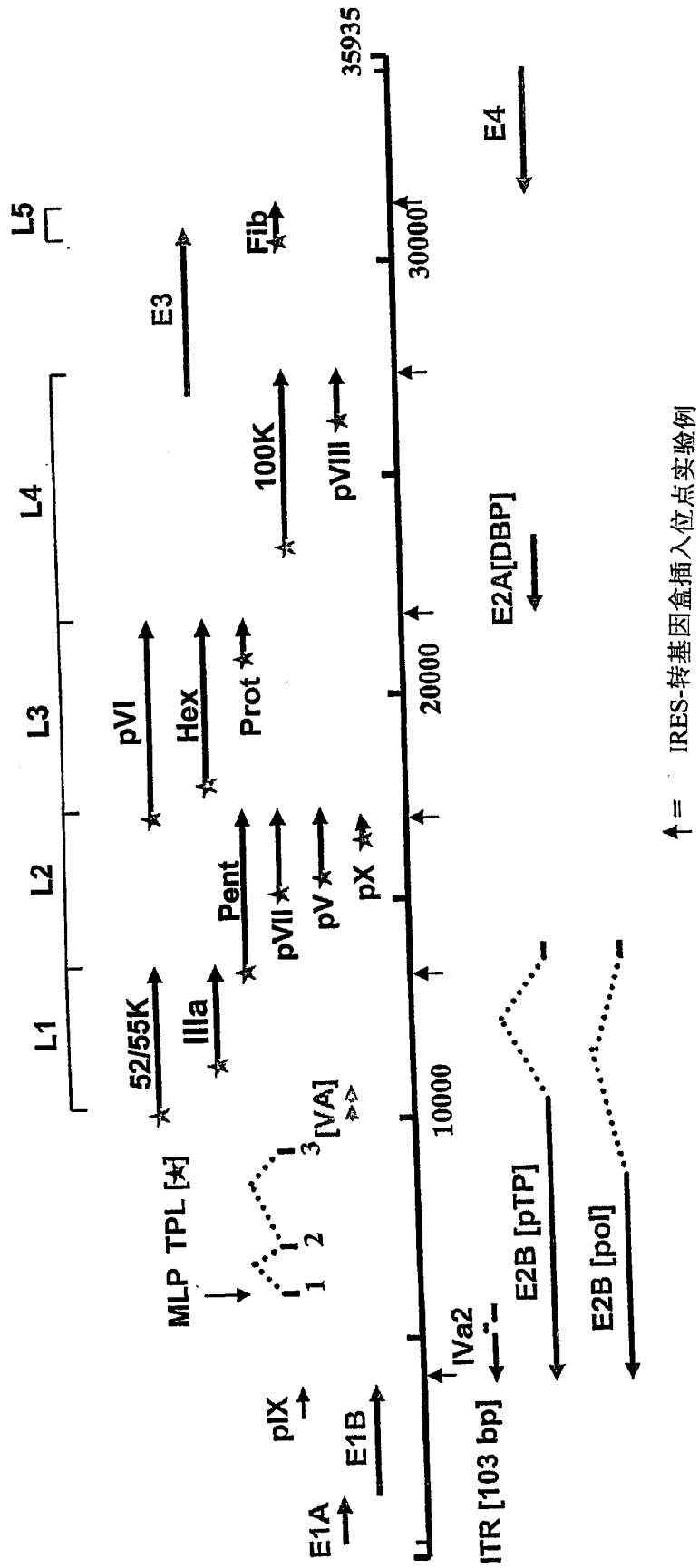
5' CT TAT TAC CGG GGT ACC CAA CTC CTC GAG TAT TT

SEQ ID NO: 61 是寡核苷酸引物 1619.144.1 的序列

5' TACAA GTA TAC GCC ACC ATG GCT ATG ATG GAG GTC CAG

SEQ ID NO: 62 是寡核苷酸引物 1619.144.3 的序列

3' TTGTA GTATAC TTA GCC AAC TAA AAA GGC CCC



↑ = IRES-转基因插入位点实验例

图 1

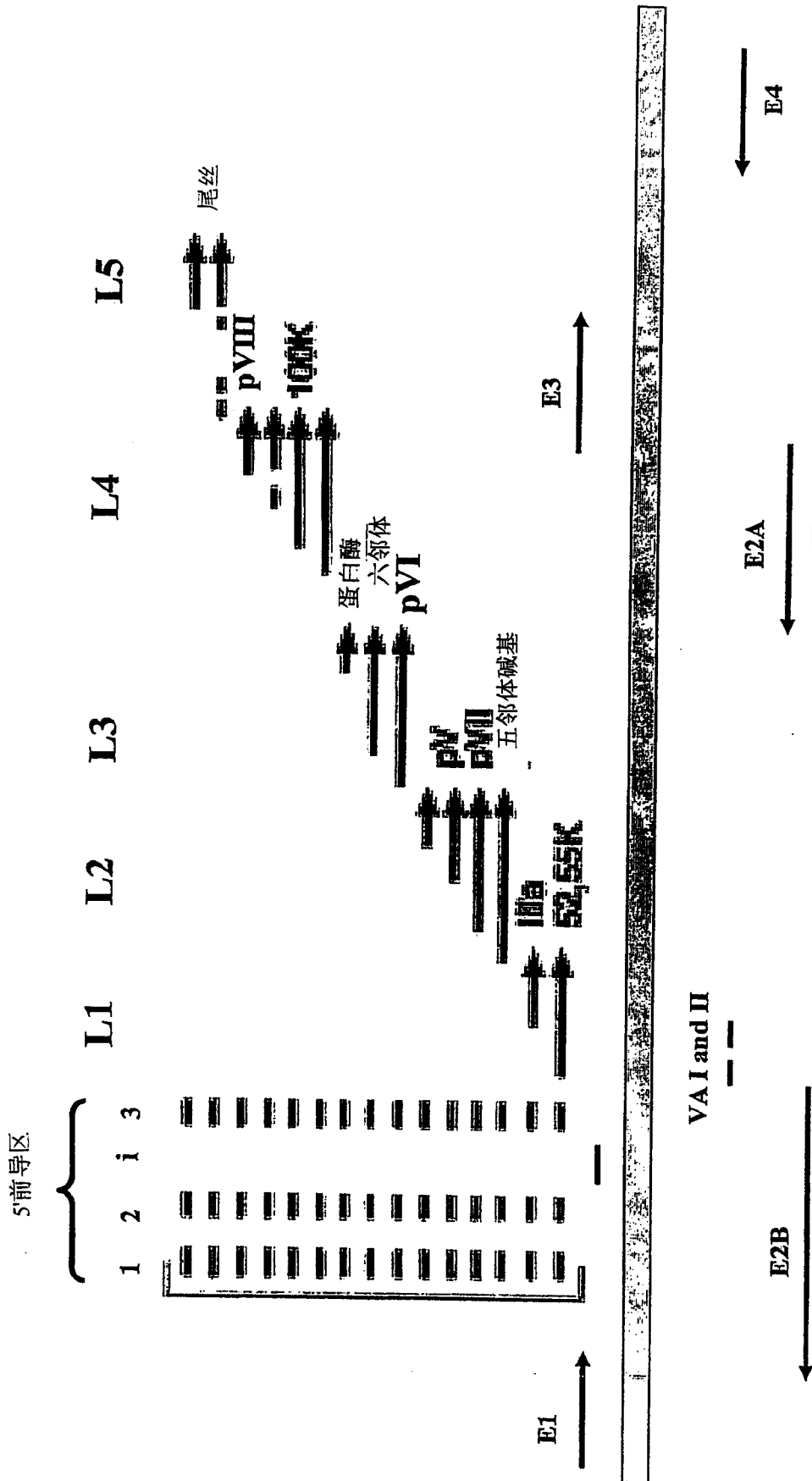


图 2

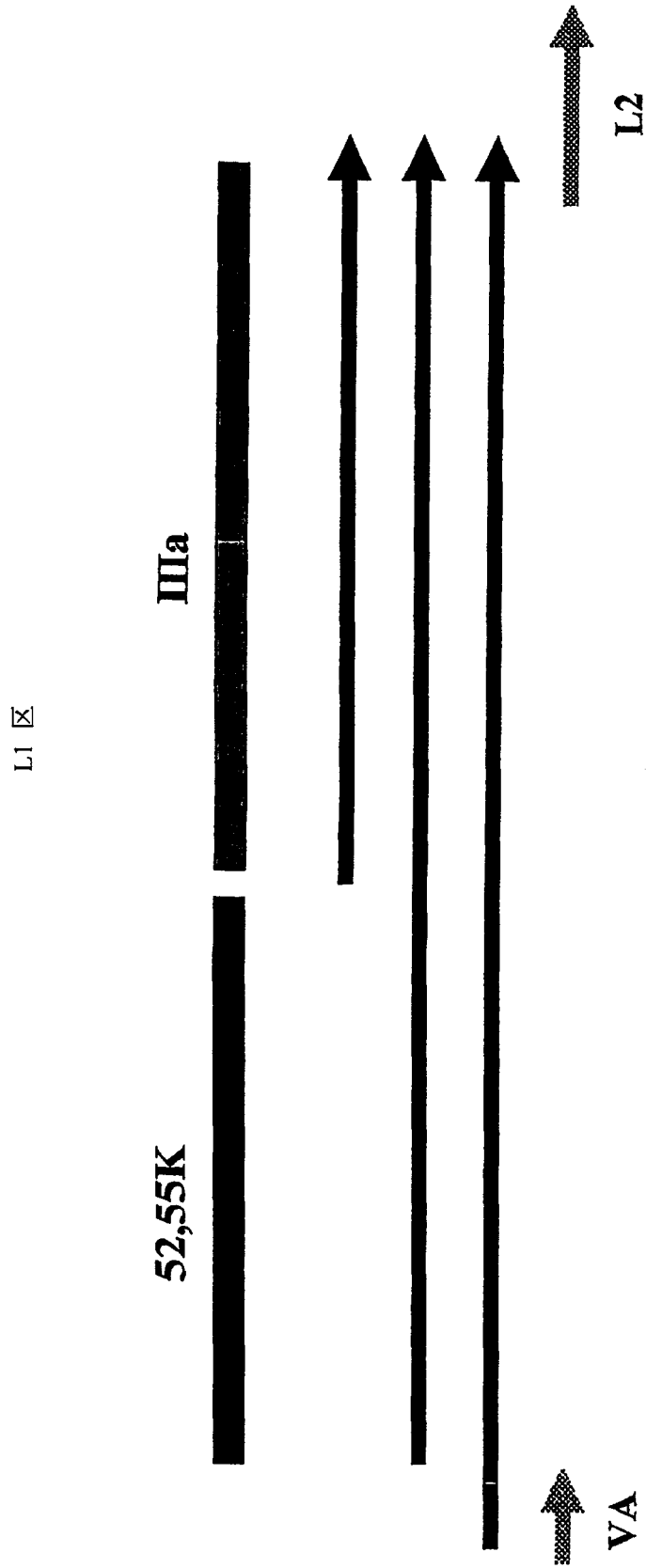


图3

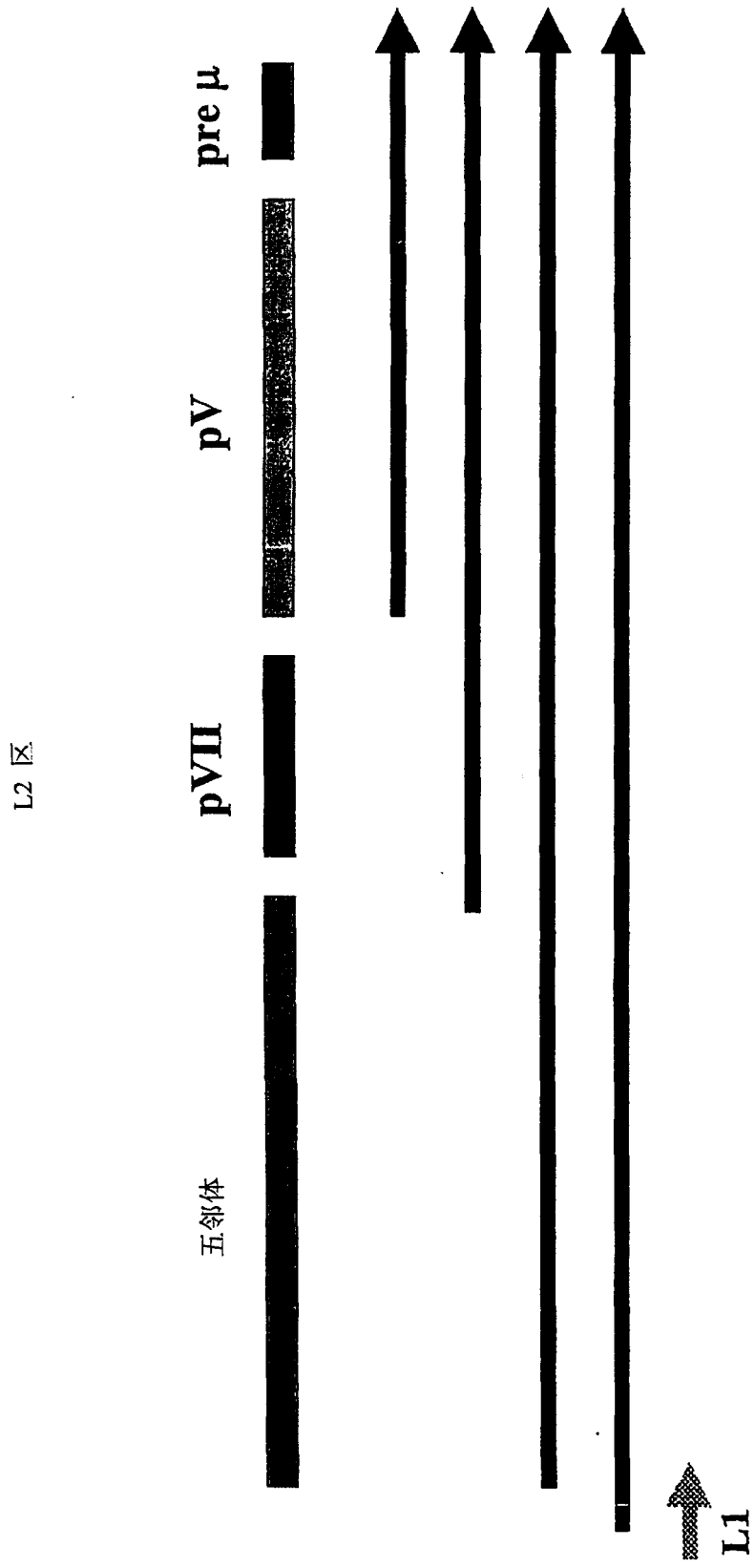


图 4

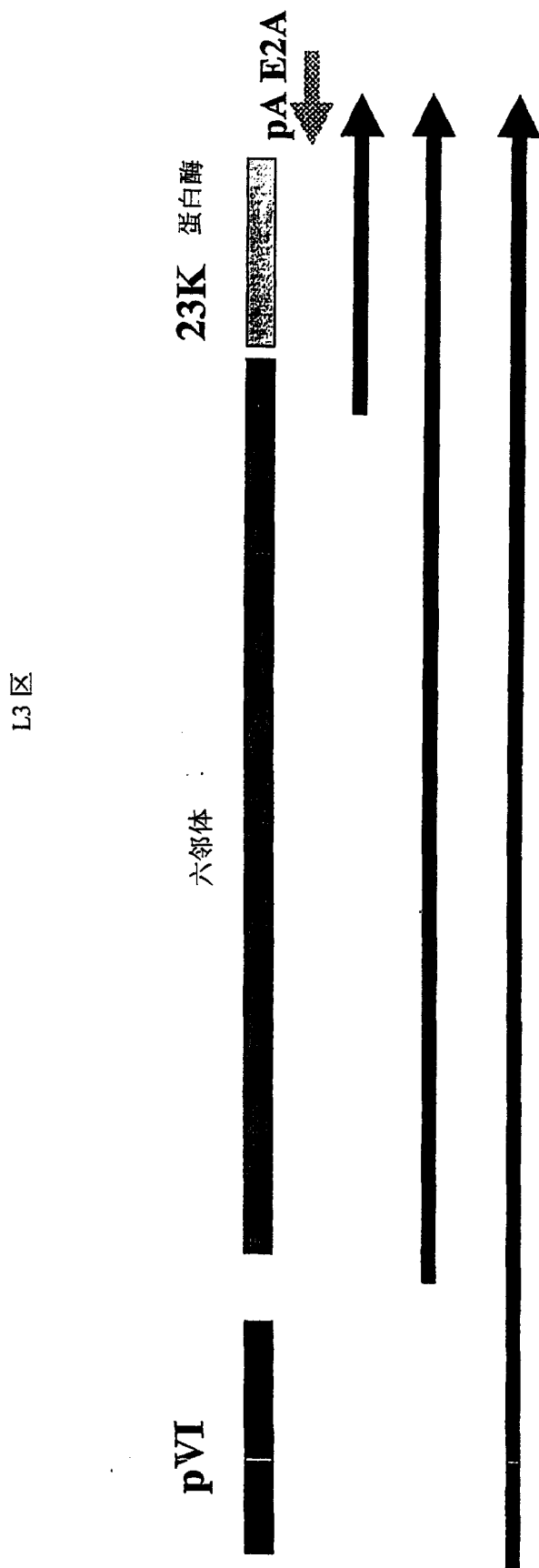


图 5

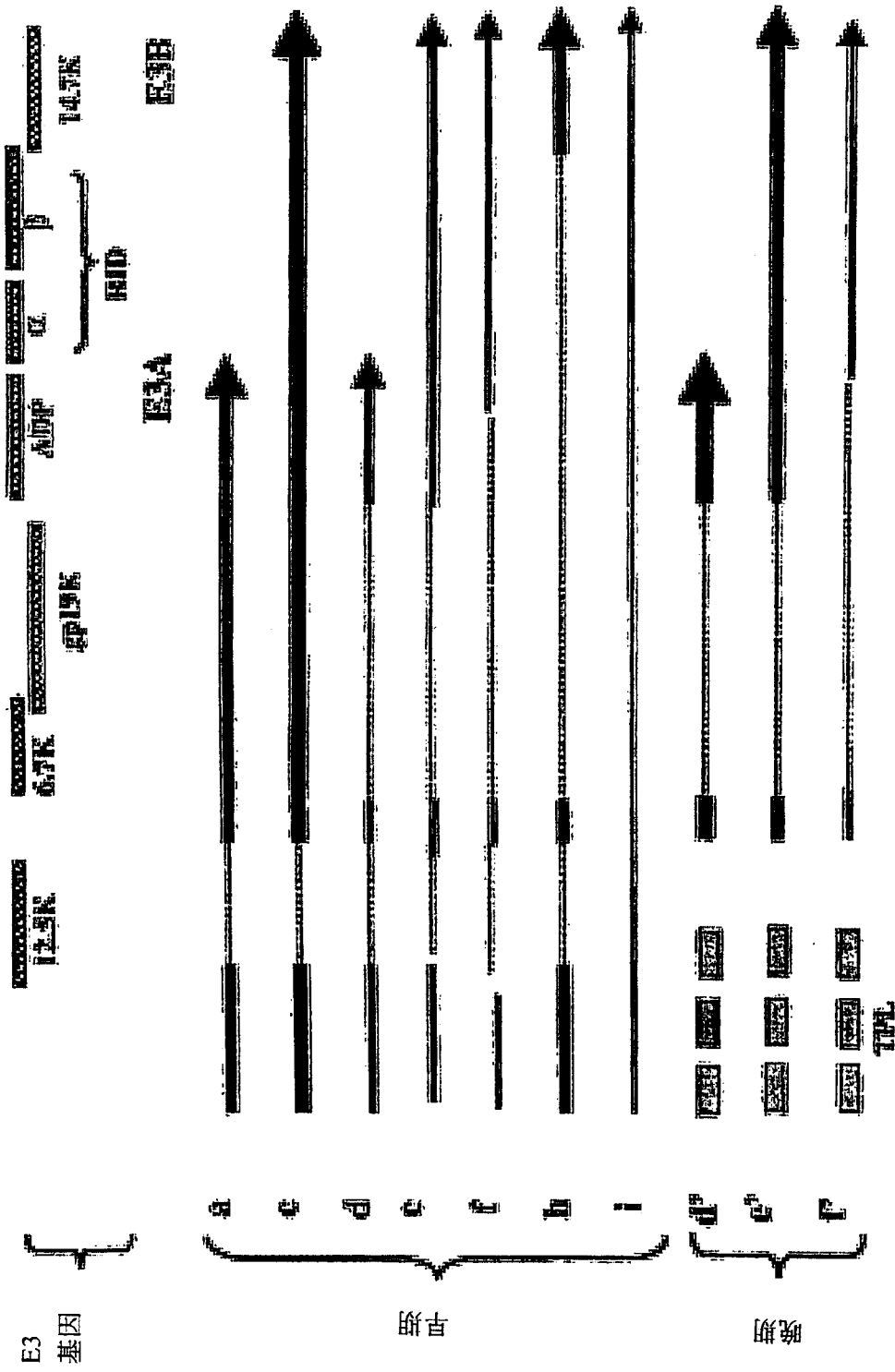
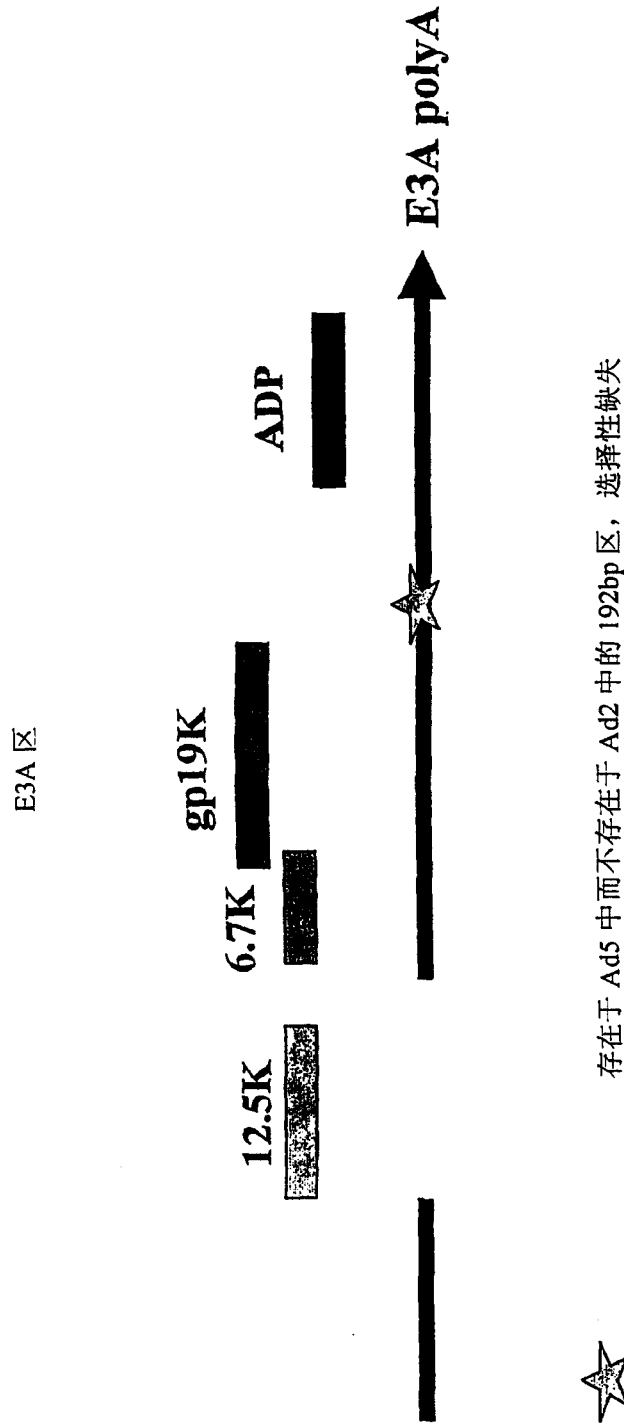


图 6



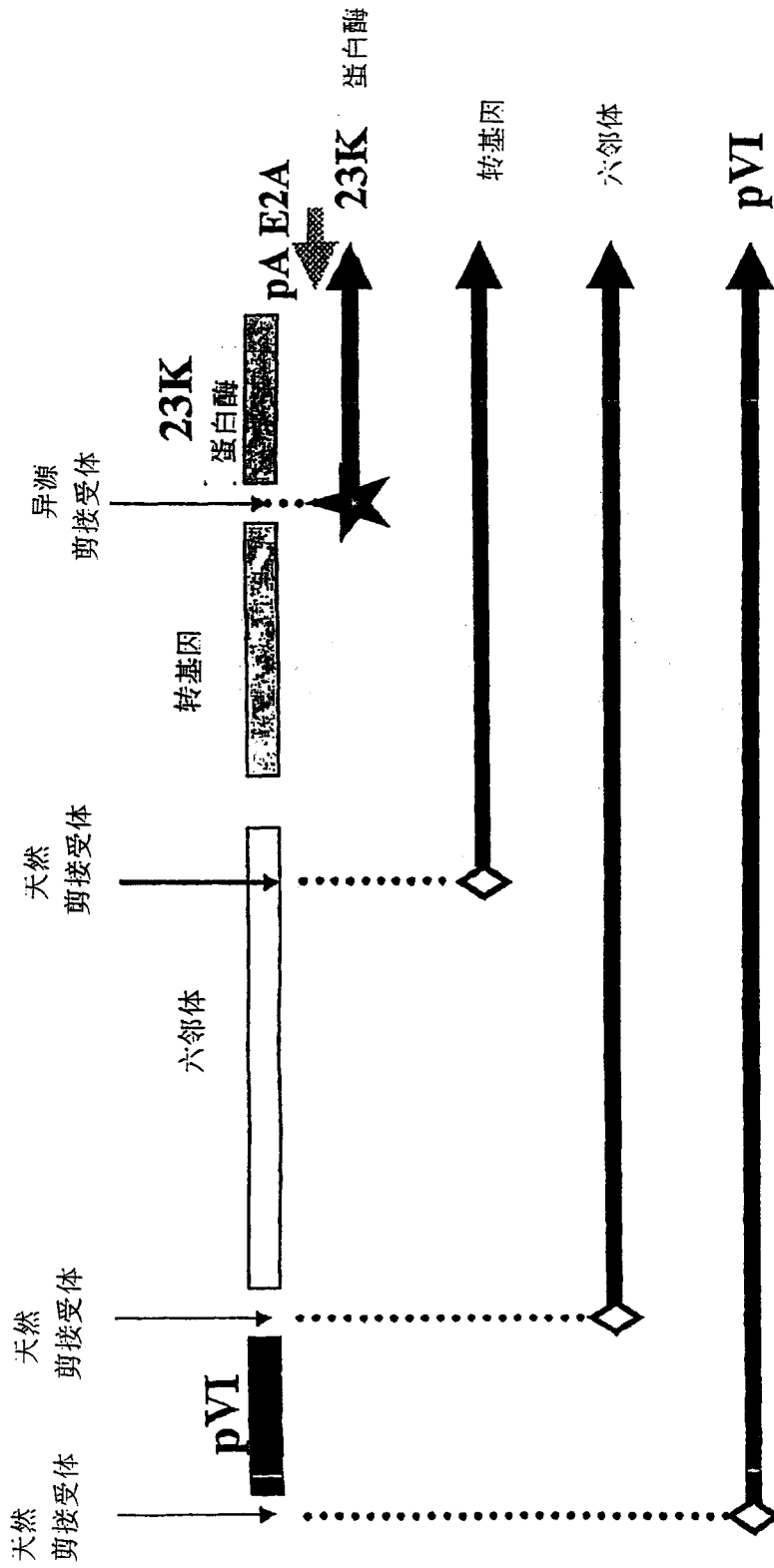
存在于 Ad5 中而不存在于 Ad2 中的 192bp 区, 选择性缺失

图 7

E3B ☒



图 8



L3 区

图 9

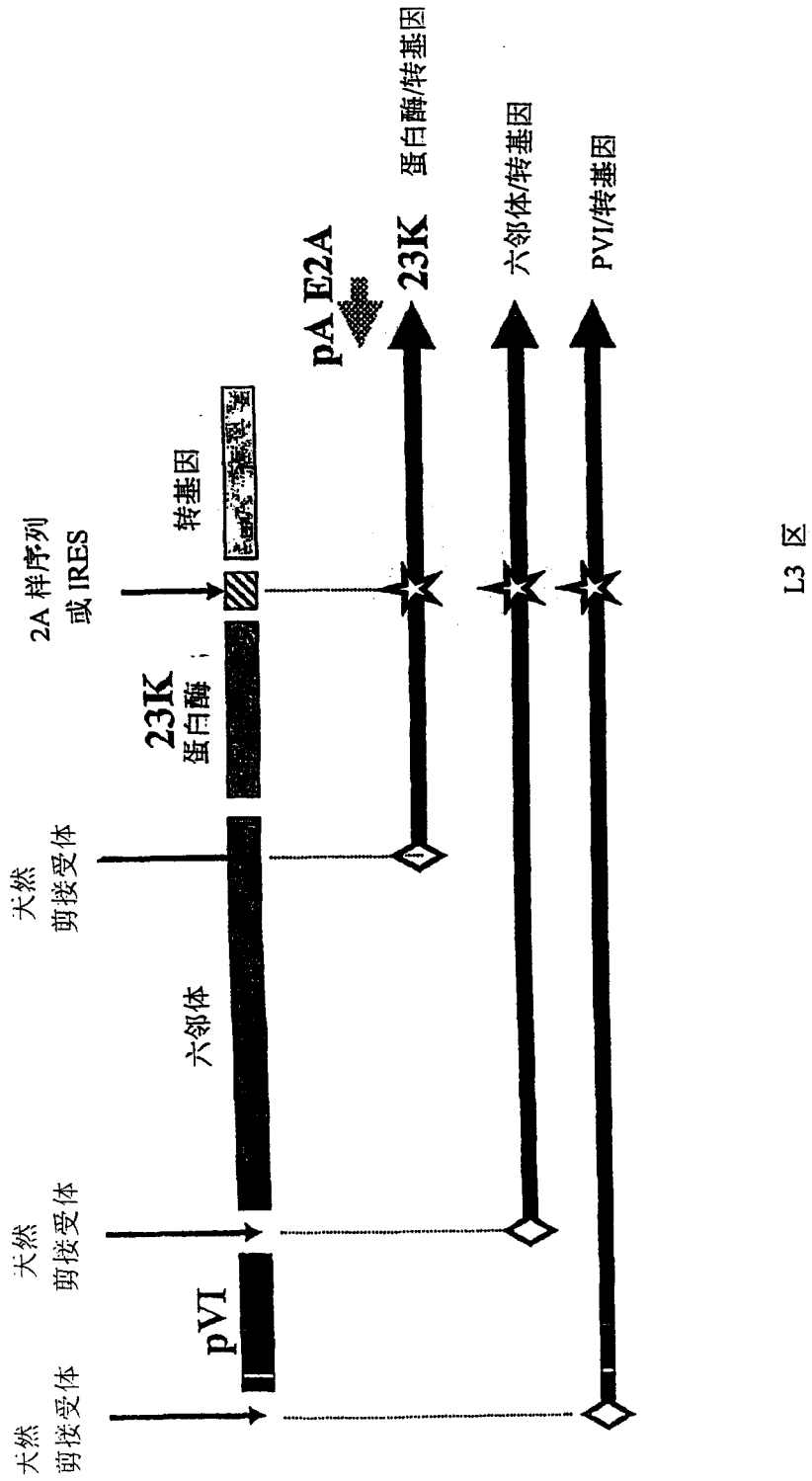


图 10

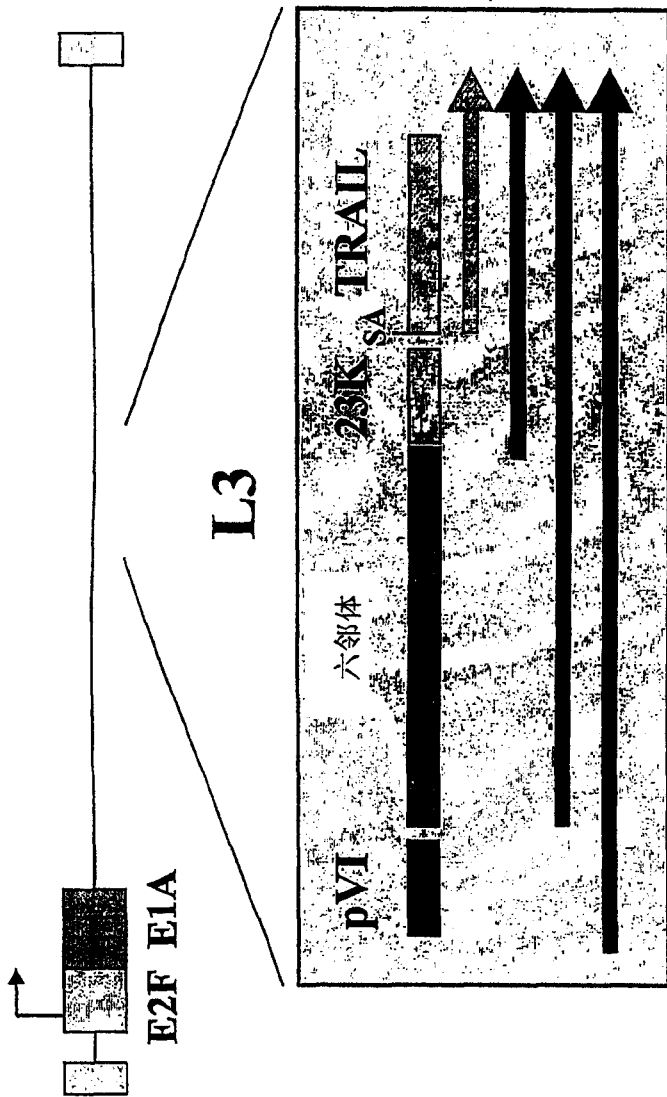


图 11

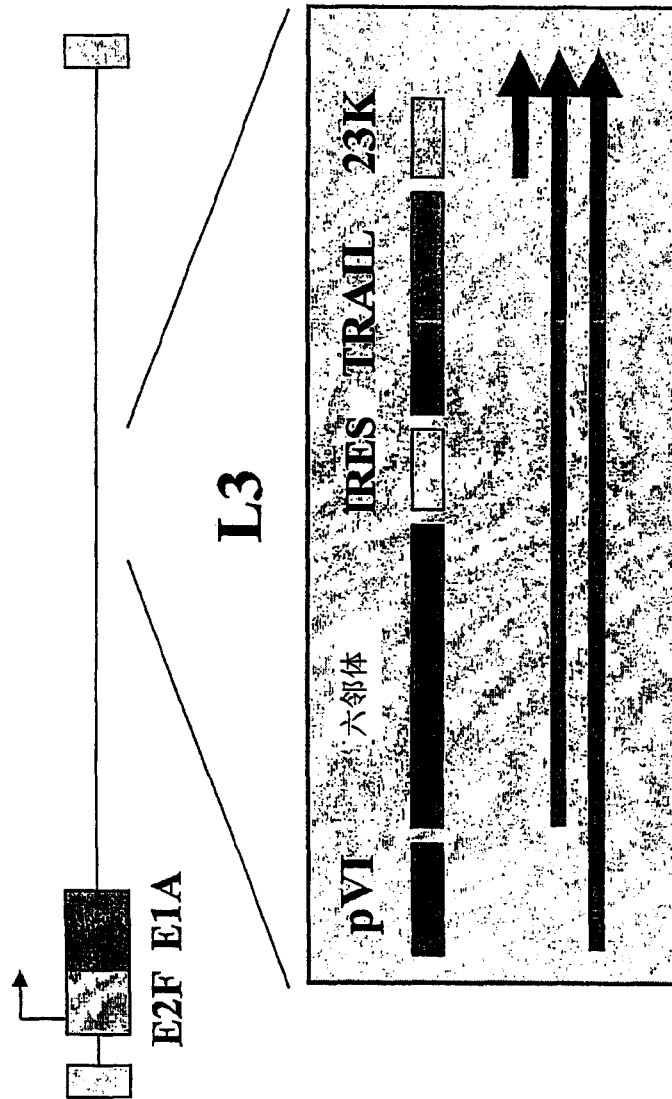


图 12